ENGLISH PROFICIENCY

IN THE SAUDI AIR ACADEMY:

VALIDATING A NEW TEST BATTERY

THESIS SUBMITTED TO THE UNIVERSITY OF EDINBURGH

by

GHURMALLAH A.J. AL-GHAMDI, B.A., M.A.

in fulfilment of the requirements of

the Degree of Doctor of Philosophy

Department of Applied Linguistics

University of Edinburgh

August, 1986

DECLARATION

This thesis is my original work and of my

own execution and authorship

Ghurmallah A.J. Al-Ghamdi, B.A., M.A.

DEDICATION

This thesis is dedicated to my parents

ACKNOWLEDGEMENTS

Many people helped to make this study both interesting and possible.

I would here like to thank my sponsoring institution, the King Faisal Air Academy in Saudi Arabia represented by its Commandant, General Mansour Al-Jawini who provided me with a three year scholarship.

I would like to thank the cadets of the K.F.A.A. who took the tests, the Officers Commanding Academic and Flying Wings, the instructors and the administrators who took an interest in the project both practically and theoretically. I would like to thank those instructors (Flying, Aerosciences and English) who helped in the administration of the tests and took part in completing the questionnaires and the rating forms.

I would like to thank my friend Abdulaziz Al-Bednah for providing me with needed results in the process of predictive validation studies.

Thanks also go to Mr. John Beach and Mr. Bryan Blyath of the Air Service Training in Perth, Scotland and to Mr. John Venables of the K.F.A.A. for providing me with language materials in the initial stages of this project.

In particular I am grateful to my supervisor, Dr. Alan Davies for his inspiring supervision. He has not only read and commented on earlier versions of this thesis but made his own work readily available.

Thanks also go to Mrs. Irene McLeod who taught me the secrets of the Spssx package which was used extensively in the analysis of the data.

Thanks also go to Mrs. Joyce Ferrington for typing this thesis.

Last but not least I would like to thank my wife and children who put up with so much during our stay in Edinburgh.

TABLE OF CONTENTS

Chapter Three
The State of the Art of Language Testing
Review of the literature

Chapter Five
Test of English for Air Cadets
Field work
Reactions and face validity

Page

Chapter Six
Test of English for Air Cadets
Internal Analyses

Chapter Seven
T.E.A.C. External Analysis
Validation Studies

LIST OF ABBREVIATIONS IN THE MAIN TEXT AND THE APPENDICES

(1) Current system: institutions, teaching and testing systems.

A.L.C	American Language Course
A.L.C.P.T	American Language Course Placement Test
D.L.I	Defence Language Institute
E.C.L	English Comprehension Level
K.F.A.A	King Faisal Air Academy
R.S.A.F	Royal Saudi Air Force
T.S.I	Technical Studies Institute

(2) The new test: parts and battery

CL1	Cloze test one
CL2	Cloze test two
CL3	Cloze test three
CL4	Cloze test four
CLT	Cloze total scores
Dic1	Dictation one
Dic2	Dictation two
GR	Grammatical structure test
Lis	Listening comprehension test (lecturette)
Sp	Speaking test (oral interview)
T.E.A.C	Test of English for Air Cadets

(3) Abbreviations of criteria used in concurrent and predictive validity

A.B.S	Ability to speak (instructors' estimates)
A.E	Aeroengine 1
A.I.E	Aeroengine 2
A.I.R	Airmanship
A.V	Average
A.V.N	Avionics
D.Y.N	Dynamics
E.C.T	English Comprehension Test
G.P.A	Grade Point Average
M.A.T.H	Mathematics
M.E.T	Meteorology
N.A.V.	Navigation
Phy	Physics
Rating T	Rating total
U.L	Understanding listening
U.R	Understanding reading

(4) Standardized Tests

C.P.M	Communicative Proficiency Measure
E.L.B.A	English Language Battery
E.L.T.S	English Language Testing Service
E.P.T.B	English Proficiency Test Battery
T.O.E.F.L	Test of English as a Foreign Language

LIST OF TABLES AND FIGURES IN THE MAIN TEXT

ABSTRACT

When the validity of an operational syllabus or a test is questioned, in relation to a particular use, then either should facilitate modifications or changes.

One of the most common complaints of instructors and administrators alike of the language programme (teaching and testing materials) at the Royal Saudi Air Force (R.S.A.F) institutions in general, and the King Faisal Air Academy (K.F.A.A) in particular, is that it prevents them from doing any useful work with cadets.

Concern about the validity of the test and the teaching materials led to the setting up of this project by the K.F.A.A in Saudi Arabia.

Its purposes were firstly, the assessment (investigation into the validity) of the test and the teaching materials produced by the Defence Language Institute in the U.S.A and used by the R.S.A.F, in relation to their use in this context, but most importantly the production of new test material to assess the English proficiency of trainee pilots who are going to pursue a course in aeroscience and flying conducted in the medium of English. The ultimate aim was to provide a prognostic test which it is hoped will improve the teaching of English by discouraging the teaching of "synonyms and encouraging the teaching of English" with emphasis on listening, speaking and reading.

Relevant work in the U.K and the U.S.A is described and discussed. The theoretical principles underlying the construction of the American Language Course (A.L.C) are described and discussed.

The shortcomings of the A.L.C and the E.C.L are presented and

discussed. Alternative specifications for a new syllabus are described and discussed.

The principles underlying the construction of the new test material, the Test of English for Air Cadets (T.E.A.C) are described and discussed and a number of hypotheses are set up. These hypotheses relate to the:

a. English Comprehension Level predictive and content validity;

b. Test of English for Air Cadets concurrent and predictive validity;

c. Construct of language proficiency;

d. T.E.A.C's discriminant power between different levels of performance.

Preliminary, pre-pilot and pilot test work is described and the final version of the T.E.A.C. is considered in detail. Test analyses of the data, both internal and external are described and tabulated. The analyses considered are item analysis, reliability of the sub-tests and the battery as a whole, test inter-correlations, factor analysis and Analysis of Variance. Concurrent and predictive validation studies are considered in some detail. Expectancy tables are used to establish cut-off points. The hypotheses are considered and on the basis of evidence they are all rejected in their null form.

As a result of these analyses it is argued that the T.E.A.C. shows satisfactory reliability and validity. One optimum cut-off is established for all academic subjects for the two levels tested.

The implications of the results are discussed and a follow-up study is recommended.

CHAPTER ONE

CONTEXT, BACKGROUND AND THE PROBLEM

This chapter concerns itself with introducing the context in which this research project was conducted. It also gives a brief historical perspective of the educational development in Saudi Arabia, the structure of the educational system and the main supervising authorities of education. It further examines briefly the language situation with particular reference to the status of Arabic and English. More importantly, it introduces the background of the problem to be examined in this thesis and reports on a validation study made on the present testing system (E.C.L.) and finally it outlines the general layout of the thesis.

1.1 Saudi Arabia: Location and significance

For a number of reasons, the significance of the Kingdom of Saudi Arabia is indisputable. First, it is strategically situated as a bridge between the western world and the continent of Asia. It is, as Al-Farsy (1978) says:

> "... With Africa on one side and Iran and south Asia on the other, it is in the middle of the strategically important Indian Ocean area which is a zone of contention between the Communist and non-communist centres of power"
>
> (Al-Farsy 1978 Introduction)

Another, but, a more important reason for the significance of Saudi Arabia, is that the country being ranked as the third largest oil producer in the world adds more economical and political weight to its position among the world.

Third, and most important is that Saudi Arabia is viewed as the holy land for millions of Muslims all over the world. These and

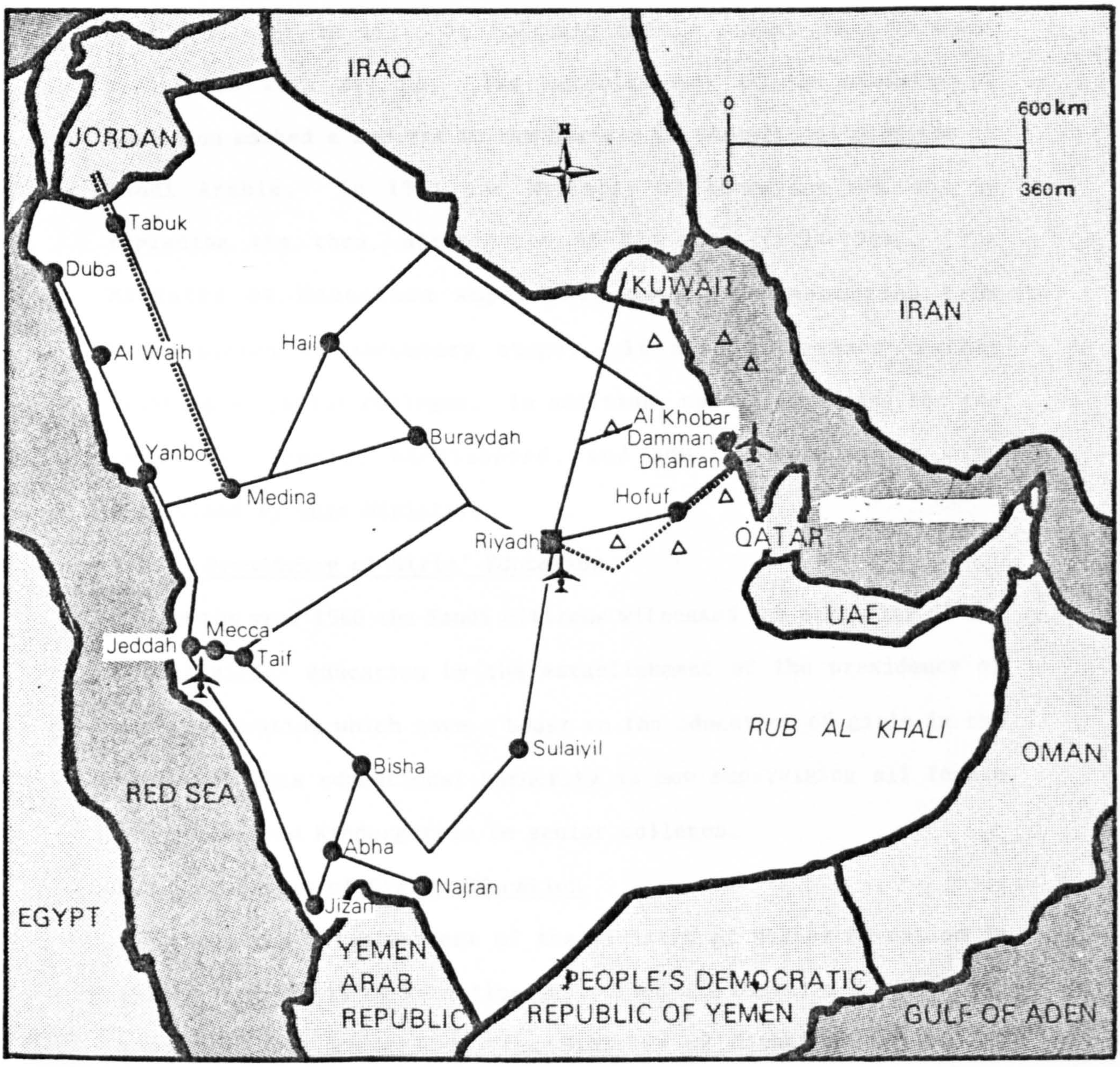
many other reasons, [for a recent case study see Al-Farsy, 1978], add to Saudi Arabia's unique position among the world.

According to the census of 1974, Saudi Arabia had a population of seven million. According to my estimation this figure has risen by about two million, since then, the majority of whom were the immigrant workforce. The country (see Fig.1 on the next page) occupies most of the Arabian peninsula and is bounded on the south by the two Yemeni Republics North and South (Aden), on the north by Jordan, Iraq and Kuwait; on the east by the Gulf, Bahrain, Qatar and the United Arab Emirates; and on the west by the Red Sea. Geographically the country is divided into four regions. The first is Najd, and is located in the middle of the country; secondly the region of Asir in the south west of the country; the third is the Hijaz, the region along the Red Sea coast, and finally the Al-Hasa region or the eastern province region located in the east of the Kingdom.

Since, in this thesis, we are concerned with both education and language, the following paragraphs are going to deal briefly with the educational system in the Kingdom of Saudi Arabia as a whole, the language situation and finally the problem to be investigated here.

FIG. 1.

SAUDI ARABIA



1.2 The Educational System in Saudi Arabia

1.2.1 Ministry of Education

With the establishment of the Saudi State by the late King Abdulaziz in 1931, it followed that a number of government ministries were set up. The establishment of the Ministry of Education marked a new era in the history of educational progress in Saudi Arabia. In 1953 the Ministry of Education was set up replacing the then, directorate of Education (1924-1953). The Ministry of Education supervises only male education from Kindergarten to secondary stage. It also supervises teacher training at Junior colleges. In addition, special training for the (male and female) handicapped, and adult education is also supervised by this Ministry.

1.2.2 Presidency of Girls' Education

In the year 1960 the Saudi citizens witnessed the beginning of the formal girls' education by the establishment of the presidency of Girls' Education which gave a boost to the education of girls in the Kingdom. This educational authority is now supervising all female education from Kindergarten to senior colleges.

1.2.3 Ministry of Higher Education

Until the establishment of the Ministry of Higher Education in 1975, the Ministry of Education supervised all higher and university education, but since then, the Ministry of Higher Education has taken the responsibility of implementing the Kingdom's policy in the field of higher education for both males and females in co-operation with the present and future universities.

1.2.4 The General Organization for Technical Education and Vocational Training (GOTVT)

At its foundation in 1980, the GOTVT was set up to supervise the technical and vocational training in order to prepare qualified personnel for different kinds of technical jobs, which were until the beginning of 1980 the responsibility of the Ministry of Education.

In addition to these educational authorities, however, there are other ministries and public organizations which provide general and specialized vocational training among them for instance is the Ministry of Defence.

1.3 The language situation

1.3.1 The Status of Arabic

Saudi Arabia is a typical "diglossic" (Ferguson 1959) situation. Bakalla (1984) distinguishes more than two varieties. He described the situation as a "continuum", ranging from dilectal Arabic to Classical Arabic with marked differences in the phonological and lexical systems. Two linguistic varieties exist side by side. The two linguistic varieties are Standard Arabic (high variety) and the Colloquial Arabic (low variety). The two varieties may be distinguised in terms of the specialization of function. Each has specific function roles. While standard Arabic is reserved for formal letters, university lectures, news broadcasts, newspaper editorials and the like, the low variety (the colloquial form) is acquired as a mother tongue, on the other hand, the high variety is acquired (or learnt) at schools through formal education. However, the sociolinguistic relationship of the two varieties now have

become closer than before, i.e. while Standard Arabic is still the predominant written and spoken language in all sorts of formal contexts etc. the low variety can also be written (national stories, friendly letters, novels) and used in some radio broadcasts and T.V. transmission of some programmes.

Standard Arabic's unique role is represented by the function it holds in the whole of the Arab world. It serves as a lingua franca.

1.3.2 The status of English

1.3.2.1 The domain of education

Where English is not the medium of instruction, the second language taught at schools and other educational institutions is English. It is taught as a compulsory subject at the various stages of education. An attempt was made to teach French side by side with English but the idea was totally abandoned in the late sixties, perhaps after a bitter experience.

The education system provides six years of elementary education, three years of intermediate, three years of secondary and three to six years at Universities, Colleges etc. Postgraduate education exists for some specialities. Some children attend Kindergartens which are privately run. All education is government aided and provided free of charge for every citizen.

English is the medium of instruction for some specialities such as Medicine, Engineering, Flying training and other related scientific disciplines.

1.3.2.2 The domain of the Media

Since the discovery of oil in 1938, the oil and petrochemical industries have brought with them hundreds of thousands of

professional expatriates to the state. Moreover the country has in the past 10 years witnessed an era of development in almost every aspect of civilisation; road construction, agricultural projects, setting up of telecommunication systems, electrical networks, introduction of scientific education, all these projects needed the help of professionals who mainly come from English speaking countries all over the world and from countries which have passed the developing stage.

All those needed to communicate with government officials but more importantly they needed to know about the State's regulations and customs and what is going on in the country and abroad; it was therefore realized that need existed for the press to publish in an additional language. The press is now published in two languages, Arabic and English. There are seven daily morning newspapers in Arabic and three in English, viz. "Saudi Gazette", "Arab News" and "Riyadh Daily".

Saudi Arabia Radio and Television Services are the responsibilities of the Ministry of Information. Radio transmits in both Arabic and English. There are also two television channels, channel 1 and 2. Channel 1 transmits its programmes in Arabic and channel 2 in English.

1.4 The teaching of foreign languages at school

The teaching of foreign languages in Saudi Arabia was inaugurated in the year 1927. English and French were the only foreign languages introduced as compulsory subjects to be taught at intermediate and secondary schools. The teaching of French was later (in the late sixties) abandoned but English was retained. The

apparent reasons for retaining English were:

A. English is the language most widely used to report advances in science, medical and technological research. Saudi students who will eventually read for degrees in science and technology need to acquire English in order to cope with their studies. Adequate proficiency in English is fundamental in order to put up with English language medium instructions as well as the large amount of references which they are required to read.

B. Studies abroad have been a significant complementary feature of higher education development for the Kingdom. English speaking countries, particularly the U.K. and the U.S.A. are the main countries in which Saudi students pursue their higher education. Adequate proficiency in English is among the requirements of admission to study in these countries. For this reason English was retained.

C. Because English has been accepted as an international language, the need arose to transmit the knowledge and sciences of the Muslims to other people in this language.

One of the main duties of the Muslims is to spread the message of Islam. English is seen to be one of the most eligible languages in the world to fulfil this purpose. The holy Quran and related subject matters have been and are still being translated into English.

1.4.1 The aims of teaching English in schools

The aims of teaching English as seen by the Ministry of Education are:

A. To produce in three years an individual who is able to speak, read, listen with understanding simple current English and to

write a connected passage of up to half a page about a simple subject or incident.

B. To give pupils who finish their formal education in the third intermediate enough knowledge of the language to help them in their vocational training.

C. To give pupils who proceed to the secondary stage a sound foundation on which to build their future studies.

D. To lay the foundation of a knowledge of English so that later they can acquire sufficient grasp of the language to enable them to preach their religion to English speakers.

1.4.2 Duration of learning and proficiency

The Saudi students, in general, spend at least 10 years learning English, among their other courses of study. English proficiency is generally inadequate when compared to the amount of time spent in learning. At schools, they learn English for six years with an average of four classes a week. At the University level, they spend three to four terms learning English with an average of two to three classes a week per term. The apparent reasons for the inadequacy of students' English might be attributed to the following reasons:

A. The diversity of teachers' background and training, along with their few numbers and instability. Saudi teachers are rarely found. The majority of English language teachers come from neighbouring Arab countries but some come from the U.K. and the U.S.A.

B. Inadequate instructions and inadequate methods of assessments. The teacher plays the dominant role in the classroom. He/she presents his/her lesson in a form of a lecture (teacher centred situation). Seldom, there has been real exchanges between him/her

and students. Assessment is largely subjective.

1.5 The problem and its extent

It was not so long after the introduction of the American Language Course (ALC) and its associate testing system American Language Course Placement Test (ALCPT) which expresses attainment in English in terms of English Comprehension Level (E.C.L.) in 1975 to replace the then ad hoc language programmes at the Royal Saudi Air Force (RSAF) training institutions, that the shortcomings of the course and above all the negative back-wash effect of the E.C.L. on language teaching have been realized.

This section concerns itself with establishing the negative back-wash effect the E.C.L. has on the teaching of English at the King Faisal Air Academy (K.F.A.A.) in Saudi Arabia. Its secondary aims are to give a brief description of the course and the test. However, a whole chapter (2) will be devoted to the present syllabus, illustrating its shortcomings, assessing its validity and outlining new specifications of what is hoped to be the appropriate English course for the cadets of the K.F.A.A. This chapter (1) also justifies the change through changing the testing system first rather than the syllabus. The remaining chapters, 3, 4, 5, 6, 7 and 8, will deal firstly with examining the State of the Art with regard to language testing and above all the design and validation of a new English language proficiency test to be used at the K.F.A.A. which it is hoped will replace the existing testing system and promote the English language teaching and testing development at the Academy.

1.5.1 Current situation

The Royal Saudi Air Force (RSAF) requires its flying and technical personnel to possess a working proficiency in the English language. English is not only the medium of instruction in the RSAF's educational institutions, but it continues to be the language used at work upon completion of training.

In order to account for the diversity of the candidates in so far as their English language proficiency is concerned, training in the English language has to be provided. Few candidates possess adequate proficiency in the English language which enables them to pursue English medium courses upon entry, therefore, an English language programme has to be provided.

Most training in both English and specialities is given by qualified native and non-native English language speakers. On the whole the native speakers represent British/English speakers and the non-natives are Saudi Arabian instructors who have received their undergraduate training in Saudi Arabia but pursued postgraduate work in either the U.K. or the U.S.A. or in both countries.

Full time intensive English language training (ELT) is provided for future pilots at the King Faisal Air Academy (KFAA) and for future aircraft technicians at the Technical Studies Institute (TSI).

The ELT system now in use in the above mentioned institutions is based exclusively on the American Language Course (ALC) produced by the United States Defense Language Institute (DLI).

The proficiency test associated with the ALC is known as American Language Course Placement Test (ALCPT), attainment in this test is known as English Comprehension Level (ECL).

1.5.2 ALC general description

The ALC has originally been devised by the DLI for the purpose of teaching English to non-American military personnel who were required to undergo various kinds of military training in the U.S.A. The ALC consists of 17 textbooks (volumes); these volumes are distributed over three phases, elementary, intermediate and advanced. Each phase contains a number of volumes and each volume carries a number. The following sequencing is provided:

The elementary phase

Volume 400
500
600
700
1200
1300
1400

The intermediate phase

Volume 2100
2200
2300
2400

The advanced phase

Volume CP1 (continuation phase 1)
CP2 (continuation phase 2)
2500
2600

1.5.3 ALC as used in the K.F.A.A.

1.5.3.1 The current situation

The main cadet intake to the Academy is completed after the month of Ramadan of each year. Each cadet first undergoes 5 weeks of initial military training and then begins English language training. In addition to the main intake cadets are admitted to the Academy throughout the year. All new cadets are given two ALCPT's. The

first ALCPT is given for familiarisation purposes of the test's format and procedures, the second is given immediately after the first one, normally the two tests are given on two successive days.

1.5.3.2 Cadets programming in the ELT

Cadets are programmed into ELT on the basis of the score obtained on the second test in accordance with the DLI programme of instructions. The chart below illustrates placements within the volume of the course in accordance with the score obtained:

Entry ECL score	ALC volume
73-74	2600
71-72	2500
68-70	CP II
65-67	CP I
59-64	2400
53-58	2300
46-52	2200
40-45	2100
38-39	1400
36-37	800
34-35	1300
32-33	1200
30-31	700
0-29)	1100
0-29)	600
0-29)	500
0-29	400

1.6 Defence Language Institute proficiency tests

1.6.1 Historical perspective

Between the years 1948-1953, proficiency tests in 31 languages were designed and used by the DLI. These tests, as was reported by Peterson and Cartier (1975) were considered inadequate in the sense that they failed to discriminate between candidates of different levels of proficiency. In the year 1954 following the discovery of the inadequacy of these tests, a new series of tests were developed.

It was not long before it was realised that these new series were not as valid and as reliable as had been assumed. In the year 1966 the DLI proposed to design a new series of tests aimed at improving the previous tests. The English version of these tests is referred to as ALCPT (E.C.L.) and is the one now in use by the K.F.A.A. and other R.S.A.F. establishments.

ALCPT (E.C.L.)

1.6.2 Design and specifications of the test

Analysis of the content of the ALCPT reveals that the design of ALCPT Battery was drawn up on the basis of certain linguistic categories to be tested. Testing through listening and reading were the prime consideration of the test.

Items of the test were based on two linguistic levels, lexis, which represents the majority of the test items, and grammar, the remaining items are listening comprehension.

The test Battery contains one hundred items that concentrate on the following points:

a. Lexical items

    1. General American/English vocabulary

    2. Idiomatic expressions

b. structural items

    1. grammatical usage

    2. word order

1.6.3 Test format

The first 60 items are listening and the last 40 are reading type items. The listening part of the test, recorded on tape, contains items which require cadets to listen to statements or dialogues on

the tape. The cadets choose from among four alternatives, the most appropriate answer.

The reading part of the ALCPT is designed to test the ability of a student to understand short written material. This portion concentrates more on grammatical usage. From among four alternatives, cadets choose the most appropriate answer.

1.6.4 *Item illustrations (lexis)

1. Ray has an expensive item.

   a. broken

   b. bad

   c. costly

   d. cheap

2. The conductor the train's departure.

   a. announced

   b. spoke

   c. said

   d. told

Grammatical structure

3. I happy to see you today.

   a. is

   b. am

   c. be

   d. very

   Stem: choose the correct sentence

4. a. To get me work it to takes 45 minutes.

   b. It takes me 45 minutes to get to work.

   c. To get to work it takes me 45 minutes.

d. 45 minutes it takes me to get to work.

* These are actual items taken from one of the ALCPT's booklets.

1.7 ALCPT (E.C.L.) and its negative back-wash effect

There is no doubt that tests/examinations affect the teaching-learning situation. By the introduction of the A.L.C and its associate testing system ALCPT, what is known to the cadets of the K.F.A.A. and other R.S.A.F's educational institutions, as the E.C.L. in 1975 to replace the, then, ad hoc language programme it was thought that problems of language teaching/testing had disappeared for good. It was not too long after, that dissatisfaction had been expressed about the course and its test. Dissatisfaction was not because the A.L.C. is out of date in the sense that it represents a more traditional view of language teaching/testing that of "structural" and that new teaching/testing methodologies, for example, "Communicative" have appeared and should have replaced the existing one. It was soon realised that the E.C.L. dictates what is taught and how.

By experience, cadets at the K.F.A.A. became familiar with the contents of the E.C.L., it is not that they memorise the actual items in the test Battery, but the type of items.

What is required of them is to obtain a pass in the E.C.L., 75 out of a maximum of one hundred. Pressing the English language instructors at the Academy for synonyms and lists of vocabulary items of the units of thc A.L.C. is the sole demand by the cadets. In actual fact the E.C.L. is dictating what is taught. What should be taught is beyond the E.C.L. capacity. Cadets do realise what English they want to be taught to them. They realise that they will

be required to understand, speak, read, and to some extent write in English, but "all", by all I mean cadets and instructors, are met with the rigidity of the E.C.L. It is not at all surprising that the following teaching/learning situation exists:

Instructor: O.K. Ali, start the next sentence please

Cadet: start same begin, commence sir

I: Forget about this now

C: Forget synonym not remember ... etc.

Or worse than that "we want synonyms sir, E.C.L. next week". In brief, because of the requirements of the E.C.L., instructors are prevented from doing any useful work with their cadets, consequently they pursue subsequent phases of training linguistically incompetent in the English language.

Although the negative washback effect of the E.C.L. on the teaching practices in the Academy is a substantial reason that justifies the change and the design of a new test, it was thought that it might be of interest to provide some normative data from existing records on the E.C.L.'s predictive validity.

To this end it was possible to obtain scores of sixty three cadets in both the E.C.L. and end of term examinations in some of the aerosciences subjects. It is worth noting that these sixty three cadets are different from the two samples used to determine the internal and external validity of the newly designed test.

1.7.1 The predictive validity of the E.C.L.

Method and procedures

1.7.2 Subjects

The sample consisted of 63 cadets, all of whom have had their

language proficiency assessed by the E.C.L. All scored 75 or above in the E.C.L. out of a maximum 100. They all underwent a unified language teaching programme (ALC). All were native speakers of Arabic. They all finished their subsequent phase of training in aerosciences conducted in the medium of English and have taken their exams in seven aerosciences subjects.

1.7.3 Criteria

The external criteria used to investigate the predictive validity of the E.C.L. were end of term exams results. The results represent marks obtained in Dynamics, Aeroengine, Avionics, Airmanship, Navigation, Meteorology and the average of all scores obtained by each individual.

1.7.4 Procedures

Pearson product moment correlation coefficients were computed to express the predictive validity.

Correlations were computed between the E.C.L. and each of the aeroscience test scores.

1.7.5 Results and discussion

Low correlations were found between the E.C.L. and each of the aeroscience test scores. These correlations range from -0.068 to 0.11. Two cases (two coefficients) represent an iverse relationship between the E.C.L. and two of the criteria, which may suggest that high scores on the predictor corresponded with low scores on the criteria and vice versa. Almost all coefficients obtained (see appendix page 81) were statistically not significant. The insignificant coefficients might be attributed to factors such as:

a. The restricted range of scores on the E.C.L. The sample

mainly consisted of those who have completed the language programme successfully, the majority of whom obtained 75.

b. The arbitrary cut off score of the E.C.L. is 75. It is the pass mark score. The mode clustered around this score.

While it is realized that language proficiency is only one variable among others, affecting achievement in subsequent training, it was expected of the E.C.L. to show a positive relationship with a value of .30 at the least. Results reveal that there is no evidence of a positive relationship between the E.C.L. and the criteria.

The question which lends itself to investigation is what should one do in this case? Changing both or at least modifying both the syllabus and the test is necessary at this stage. Given the choice, what should be changed first? Should one start with the syllabus, this is the normal process, it is desirable, or should one start from the goal. A sensible and a quicker solution to the dilemma is what Davies wrote:

> "Change in language teaching must be possible; that is there must be some way of responding to new ideas and demands. It is best if the change comes in through the syllabus and the examination and the teacher. If a choice has to be made among these in order to move quickly, then undoubtedly the test/examination is the most sensitive; it is the most controllable, it acts overall, it is the most difficult (pace W. Africa) to ignore, it has most certainty in terms of its goals. The test/examination is a major and a creative influence for change and development in language teaching, and if there is a need to choose, then that is what should always change first."
>
> (Davies, 1985:p.7)

1.8 The general layout

The choice has therefore been made to begin with the testing

system. The following chapters (except chapter 2), deal with the development which has occurred in the language testing literature over the years. But the main theme of this thesis is the construction and the validation of a Proficiency language test, with which is hoped to change for the better, a long felt need for a proper language teaching and testing programme. Chapter two concerns itself with assessing the validity of the present syllabus (A.L.C.) in relation to its uses at the K.F.A.A. as well as current thinking in language teaching and linguistics. Chapter three provides a critical review of the state of the art. It examines three eras of language testing development, the three well known phases identified by Spolsky (1975, 1978) are discussed and exemplified. The notion of proficiency is considered in some detail. Traditional and current aspects of assessing proficiency are discussed and exemplified. Content and the principles underlying the construction of the T.O.E.F.L., the E.P.T.B., the E.L.B.A., and the E.L.T.S. are also reviewed. Some validation studies made on these tests are also quoted and discussed. The concept of validity is discussed in relation to language proficiency testing. Chapter four presents the concepts underlying the Test of English for Air Cadets, T.E.A.C. It reports on the test design, its content and methods. It also gives a brief item illustration. It also attempts to provide a definition of proficiency in testing in relation to the use of the test within the domain of the Academy. Chapter five concerns itself with the trialling programme of the T.E.A.C., it first describes the journey to Saudi Arabia to conduct the field work. It outlines the characteristics of the samples on

whom the T.E.A.C. was tried. It presents the T.E.A.C's consumers reactions to the new measure and finally it reports on the trialling of the measure in accordance with the aims of the pilot study illustrated in the same chapter. Chapter six reports on the internal validity of the T.E.A.C., i.e. item analysis and test reliability. It also reports on the statistical properties of the T.E.A.C as was tried on the two populations. Results of the test's intercorrelations, factor analysis and analysis of variance are presented and discussed. Chapter seven concerns itself with the validation studies made on the T.E.A.C. Results of concurrent and predictive validity are presented and discussed. The final chapter, chapter eight, presents a summary of the content of this research project and reports on the conclusions arrived at in this study. In addition, it presents two-fold implications. Firstly, pedagogic implications with regard to the shape of the language programme that should be taught at the K.F.A.A. Secondly, practical and theoretical implications with regard to the future of the T.E.A.C. and the result of its implementation.

CHAPTER TWO

ENGLISH LANGUAGE SYLLABUS AT THE K.F.A.A.

OLD AND NEW

2.1 General Layout

This chapter concerns itself with two objectives. Firstly, it thoroughly examines the content of the ALC and assesses its validity in relation to its use at the K.F.A.A. It also illustrates its shortcomings and assesses the theoretical principles underlying its construct in relation to current thinking in language and linguistics, syllabus design and language teaching methodologies. Secondly, it sets out to propose an alternative set of specifications of a new syllabus in view of experience gained over the few years during which I worked as an instructor and later as a director of the English Language Teaching programme at the K.F.A.A., the specifications are partly based on a questionnaire distributed among aeroscience, flying and English language instructors and partly on observation and personal experience.

2.2 Introduction

In the discussion which followed the presentation of a paper entitled 'Some theoretical problems and practical solutions in proficiency test validity' presented by Petersen and Cartier (1975), B. Spolsky asked:

> 'Let's say suddenly you realise that you would like to change the syllabus. How will you justify the fact that the new syllabus produces more?'
>
> (p.116)

To this, Cartier gave three reasons. The first is expressing a concern regarding its validity. The second is the discovery of an

improved methodology. The third is a change of emphasis in the teaching of the language skills. He went on to say:

> 'If you changed the content of the syllabus it would be for some logical reason and that logical reason would be just as applicable to the testing programme as it is to the course design'.
> (p.117)

Full time intensive English Teaching Training (ELT) is provided for the cadets at King Faisal Air Academy, Riyadh, Saudi Arabia. The ELT system now in use is entirely based on the American Language Course (ALC) produced by the United States Defence Language Institute (DLI). For reasons which will be discussed below, the above mentioned course is not entirely satisfactory as an English Language Course for the Academy's cadets. This chapter attempts to outline the shortcomings of the syllabus now in use, and propose a set of specifications of what the English Language Course at the Academy should be. Subsequent chapters are going to deal with the construction and validation of an English language proficiency test that should ideally accompany the proposed language programme.

My point of departure is an examination of the present syllabus in terms of its content, purpose and objectives.

2.3 Purpose and theoretical principles underlying the construction of the ALC

My analysis of the content of the ALC reveals that the course was designed with two main purposes in mind. The first is to present American English teaching materials suitable for an intensive, audiolingual and structural presentation. The second is to give the learner an intensive cultural orientation to America. So far as the

first goal is concerned, the materials are presented in accordance with certain theoretical views on language, language learning and teaching. The theoretical views expressed in the course are linguistically structural in orientation, psychologically behaviouristic, audiolingual in approach. The second goal is derived from the view which sees language as 'Culture-bound' therefore the language system is realised in topics and situations related to the cultural aspect of the target language.

Below we will assess these two goals in relation to our situation at the K.F.A.A. and find out to what extent the present syllabus is relevant (or not) to our learners' needs and to the demands imposed by their course of study, i.e. Aerosciences and Flying Training; this is one aspect of the matter. The second is an assessment of the present syllabus in relation to the latest theoretical developments in language and language learning expressed in the fields of sociolinguistics, discourse analysis and register analysis.

2.3.1 Content of the ALC

Topics and situations

A survey of the content of the units in the ALC shows that the following topics are presented. Although we are not intending to list all the topics in the ALC, it will suffice to give some examples for the purpose of discussion. Among these topics are the following:

Travel in the U.S.A.; the Constitution of the U.S.A.; dining in the U.S.A.; customs in the U.S.A.; a typical day in an American high school; the American cowboy; the American Indian; government in the

U.S.A. etc.

While we agree with Hudson (1980) when he wrote:

> 'Most of the language is contained within culture so it would not be far from the truth to say that a society's language is an aspect of its culture'.
>
> (p.83)

we will discuss this from a pedagogical point of view..

It is obvious that the above cited entries indicate an American culture bias. The bias is of use for the cultural orientation of foreign students studying in the U.S.A. but in Saudi Arabia this cultural bias is an unnecessary burden for both cadets and instructors. The pilot-trainee at K.F.A.A. is <u>not integratively</u> motivated by interest in American life-styles, since the course includes many references to aspects of American culture and social life designed to help overseas trainees to cope with the problems of learning and living in the U.S.A.; the only motivation he has to learn the language is <u>instrumental</u>, the desire to get at his chosen field of study. This point leads us to the consideration of two affective factors in second language learning. These are <u>relevance</u> and <u>motivation</u>.

2.4 <u>Relevance and Motivation</u>

According to most psychologists, a theory of human motivation is defined as 'those factors that energise behaviour and give it direction', Van Els et al (1984) quoting Hilgard (1979). In the course of second language learning, Gardner and Lambert (1972) identified two types of motivation. The first is instrumental and the second is integrative. In this discussion, we will be mainly concerned with the instrumental type of motivation since it relates

directly to our present situation. A learner of a language is said to be instrumentally motivated when the purpose of learning reflect the more 'utilitarian' (Mackay, 1978) value of linguistic achievement. In our situation the language should be learned as a means in itself, it should play an auxiliary role.

It is a fact that the students in question have on entry, a low level of competence in English. They also realise that the Academy is English medium, so they cannot fail to be aware that they badly need further language study before they embark on their field of study. In other words, they are highly motivated.

Once the students start learning English through the ALC material, the question which lends itself to investigation is: whether the high motivation the students possessed upon entry is still as high, or has it been affected, i.e. declined? For I believe that motivation decreases and increases as a result of other external factors whether they be linguistic or extra-linguistic. By linguistic factors I mean the selection of language materials, the choice of appropriate language skills to be taught and the choice of suitable methodology. This leads us to the consideration of the notion of relevance which I would like to address here as validity in relation to the selection of the ALC materials.

2.5 Validity of the Syllabus

The theoretical contribution of disciplines such as Sociolinguistics, Discourse Analysis and Register Analysis have focused attention on language use, language users and language needs and purposes to play certain roles in different situations of second language learning. With the growth of the above cited disciplines,

face and construct validity are seen to be very important aspects in considering both how one is to design a language syllabus as well as why one should do so. A syllabus which does not take into account the needs of the learners and subsequently the appropriate choice of the language materials and methodology might be described as invalid. Within the capacity of the ALC the ultimate aim is to teach the system of the language, i.e. language as 'usage' like any other structural syllabus which has served students and teachers alike for several generations (Yalden, 1983). The construction of the ALC was based on certain theoretical views on language and certain techniques on syllabus design which in view of advances in sociolinguistic theories and related disciplines have become 'invalid' or better described as out of date. It is actually the choice or the adaptation of the programme as a language training one which is considered inadequate. The Academy realises the needs of the students and the purposes to which they are going to put the language, but the selection of the programme was not satisfactory, this selection may have been justified then owing to the unavailability of a programme which would cater for several levels of competence as well as a programme of this length (64 week programme).

2.6 Usage and Use

With the theoretical contribution of disciplines such as sociolinguistics, register analysis and discourse analysis, the shift of emphasis from teaching a second language, to use Widdowson's (1978) term, 'as a stock of usage' to teaching a second language as communication has been realised in the selection of

materials for the language programmes. This implies that the language units and items should be selected on the basis of 'having a high potential value which can be realised to perform the kinds of acts of communication which the learner will have to deal with'. Below we are going to discuss 'use and usage' as two aspects of performance and as realised in teaching materials.

One demonstrates the ability of 'usage' i.e. knowledge of the grammar of the language when he is able to construct grammatically correct sentences regardless of its context. Thus, a learner who is capable of producing the following sentences in English:

I went to school yesterday

My father goes to college every day

I am walking now

The grass is cut every day

demonstrates a knowledge of part of the system of the English language, since they all denote instances of 'usage'.

Language materials organised in accordance with the structural viewpoint of language and language teaching are dominated by 'usage' as one aspect of language competence. Thus, in a classroom presentation based on structural syllabuses, the following is a familiar pattern of presentation:

Teacher: Orange

Pupil(s): I am eating an orange

Teacher: Apple

Pupil(s): I am eating an apple

The learner is called upon to repeat a sentence pattern by using different entries. What is actually taking place is a series of

.

.

responses which could be described as manipulating the sentence pattern at the syntactic level. The learner is not communicating with the teacher in a normal sense. He is only trying to acquire a particular sentence pattern to have access to the system of the language.

However, languages play a communicative role in people's lives. Teaching the language should not be confined to teaching the system of the language. On the whole, structural syllabuses tend to focus on teaching language as 'usage', but what is actually needed is a focus on teaching language as 'use' without de-emphasising the main aspect of performance which is 'usage'. In my opinion the two aspects should not be separated. Using the language effectively in communication implies a knowledge of the system of the language, i.e. 'use' includes 'usage'. While the ultimate aim in teaching a language should be communication, this may not be attained without acquiring the system of the language particularly in the early stages of language learning (for a detailed discussion, see Widdowson 1978 and 1983).

2.7 Language materials in the ALC and motivation

The selection of language materials in the ALC is made within the framework of FLT for general purposes, while the teaching of language should be directed at certain specific situations of the kinds of acts which the learner will encounter. Register analysis (Halliday et al. 1964), helps identify what would be relevant to the students' immediate and target goals at both the syntactic and semantic levels.

The selection of the appropriate language materials would help

preserve the motivation aspect throughout the course of learning, since it is not just that learners do have a high motivation for learning upon entry, but also the extent to which they see themselves learning a course as the type of course they need, in terms of the language materials as well as the language skills.

2.8 The lexical corpus of the ALC

The ALC is a general language course and is about general topics. Lexical items are again likely to be irrelevant to the students because general subject matter must generate general vocabulary. In a research project entitled 'A word count and contrastive frequency level rating of the ALC', it was found that the course is heavily weighted towards the acquisition of lexical items. A total of 7,710 lexical items was observed, in addition to 375 special expression items (Lockheed Sept 1981).

The different study units of the course are intended to be covered in approximately the same time. However, the material to be studied in these units varies considerably both in quantity and difficulty. This is particularly noticeable in Volume 1100. From the pedagogical and methodological viewpoint, the introduction of 1,266 new words in this volume has too much front lexical loading. Any cadet who does not thoroughly master volume 1100, will have lexical problems in the remaining volumes while struggling to master the remaining 6,444 words.

Language laboratory tapes are unimaginative and extremely boring for the students. They rely heavily on simple repetition, and contain few exercises designed to involve the student in thinking and manipulating language structures. Drills should aim at problem

solving rather than habit formation.

Almost every unit of the ALC presents the student with a large amount of new material, making minimal reference to what has been previously learned, i.e. linear type of programme. However, a cyclic approach is preferred to the linear, since cadets are not all expected to reach the same specific objectives of each unit at the same time but should be allowed to progress at their own rate.

So far we have been discussing the content of the ALC in relation to its content, objectives and purpose. We related that to two important factors in second language learning, these were relevance and motivation.

We will now consider the rationale behind the emergence of the ALC as a structural syllabus; this would involve a consideration of the linguistic background during which this kind of syllabus was designed as well as an appraisal of the teaching and learning strategies employed in the course.

2.9 Structural Syllabuses

2.9.1 The Linguistic Background

By the early 1940's, linguistics was viewed as an important aspect in language teaching theory. Stern (1983) reported that the Linguistic Society of America undertook to turn their experience in language description to the task of 'linguistic analysis of each language to be taught' (several structural linguists such as Hockett, Bloomfield, Moulton and many others, undertook the preparation of learning materials based on the analysis of languages to be taught for the American Army). One cannot fail to notice that because linguistics 'then' has carried an emphasis on language

structure, the teaching materials developed were influenced by these trends in linguistic theory.

Stern (1983) reported that the set of techniques employed by structural linguists in devising teaching materials were:

a. a structural analysis of the language forming the analysis of graded materials;

b. presentation of analysis by a trained linguist;

c. several hours of drill per day with the help of a native speaker and in small classes; and

d. emphasis on speaking as the first objective.

2.10 Weaknesses and Strength of the Structural Syllabus

In the structural syllabus the emphasis is on teaching the language system. Form and meaning are assumned to be on a one-to-one relation (Yalden, 1983). Meaning is to be learned together with grammatical forms. Meaning, apart from purely lexical entries, is thought to be self-evident. The argument against the structural syllabus is that the learner has mastered the system of the language, he is able to form grammatically correct sentences but what is missing is the ability to use language in an appropriate social context. This claim implies a change in the goal of teaching the language and in the view in which language is to be defined. Despite this argument I do not believe in abandoning the structural syllabus totally, indeed, learning the language system could serve as the basis for communication. Communicative competence is something which involves 'systemic competence' (Johnson, 1982), i.e. the ability to form grammatically correct sentences. A learner who does not master the grammatical rules of the system of the language

might obscure the comprehension of the message he intends to convey. So there is no harm if the structural progression is realised in a communicative framework, for I believe learning 'usage' serves as a prerequisite for learning 'use'.

2.11 The Audiolingual Theory

2.11.1 Characteristics

The ALC is audiolingual in approach. The audiolingual theory serves as the basis for the teaching-learning strategies of the ALC. The theory has been thoroughly described and criticised by Rivers (1964). This method of teaching has several special features:

a. The teaching of the four skills in the following order hearing before speaking; speaking before reading; reading before writing.

b. The separation of the four language skills.

c. The use of dialogues as the chief means of presenting the language.

d. Emphasis on mimicry, memorisation and pattern drills as learning strategies.

e. The use of the language laboratory.

f. Establishing a linguistic and psychological theory as a basis for the teaching methods

2.11.2 Theoretical basis

The audiolingual theory reflects the descriptive structural and contrastive linguistics of the fifties and sixties. Following Skinner, its psychology is behaviouristic. Language learning within the theory has been interpreted in terms of stimulus and response.

2.11.3 Appraisal

When the audiolingual theory was first described and used in the early 1960's, it raised hopes of 'ushering in a golden age' of language learning (Stern, 1983). Stern (1983) wrote:

> 'By the end of the decade ..., its theoretical basis was found to be weak'
>
> (p.465)

In practical terms, its hopes had not been fulfilled. Lack of effectiveness of its techniques was observed by teachers using it. Students it was found, got bored. While the theory claims it does not neglect reading and writing, in practice, the teacher finds himself drifting in teaching listening and speaking at the expense of reading and writing (for an assessment, see Rivers, 1964).

However, in view of these criticisms, it should be remembered that the audiolingual theory was among the first language theories to recommend the development of a language teaching theory on linguistic and psychological grounds.

Having outlined the shortcomings of the existing syllabus in terms of content, objectives and in terms of construct, we will now consider the issue of syllabus design with reference to our target group.

2.12 Syllabus Design

2.12.1 Introduction

A language syllabus may be designed on two grounds. The first involves consideration of a number of extra-linguistic elements. The second involves the selection for the linguistic content to be included in the syllabus and associated procedures such as gradation and presentation of material.

As far as the extra-linguistic factors are concerned, they have to do with the educational environment in which the course will be taught, the characteristics of the target learners as well as the circumstances in which the educational institution works.

The extra-linguistic factors can be arrived at through what has become known as 'needs analysis'. It might be appropriate here to include a critical review of 'needs analysis' before we start analysing the needs of our own cadets.

2.13 Needs Analysis, a critical review

I will begin my discussion of the needs analysis technique by quoting Willis (1981); he wrote:

> 'Nowadays needs analysis is almost a discipline in its own right and no self-respecting TEFL Course can be without it'
>
> (p.15)

He then warned that needs analysis should not be viewed as an end in itself, rather the first step in a lengthy process.

Yalden (1983) also wrote:

> 'The process is now well enough established and the chief danger has been identified as that of isolating needs analysis from the other components in language programme design'
>
> (p.102)

The above cited quotations imply a warning of some kind. This warning has been established as a result of the fact that 'needs analysis' is not in a position to provide complete answers for all problems in the process of syllabus design, so it has to be viewed as a step among others in that process. Needs analysis does not tell us how to organise the functions and concepts which when identified and turned into language forms, the question of

organising them into a coherent teaching sequence remains unsolved.

Several schemes for defining the students needs have been proposed by writers in the field of language teaching, e.g. Van Ek (1975); Richtritch and Chanceral (1980). Although Munby's book (1978) was entitled 'Communicative Syllabus Design', it might be classified as a book on needs analysis, since the major theme of the book is about the identification of communicative needs of learners which has, later in the book, been expanded by the parameters involved in identifying those needs.

The rhetorical title of the book as it stands does not take account of other variables in the process of syllabus design such as the methodological variable for instance. It is only in the epilogue of the book that Munby realises that there are other significant variables involved in the syllabus design process which he did not take account of.

Davies (1981) warned that the model might be taken as 'a model that works not a model to speculate about'. His warning came true because not only have we seen an influence of this model on the designing of EFL textbooks recently but on the area of language testing as well. We saw B.J. Carroll (1978, 1980) using the full needs specification proposed by Munby for designing the English Language Testing Service (ELTS), but later (1983) he admits:

> 'What happened was that we learnt by experience not to bother with some areas of Munby's specification, particularly the areas of attitudinal tone and micro-function, which together, he refers to as Units of meaning'.
> (p.97)

More recently McDonough (1984) wrote: 'History has already shown us

one impressive looking apparently rigorous (she was referring to Munby, 1978) but ultimately unworkable model of needs specification' Nevertheless, a major advantage of needs analysis is given by Cunningsworth (1983) when he says: '... to free the language teachers and syllabus designers from complete dependence on their own intuition and received wisdom'

However, in the context of language teaching and testing the prime consideration of needs analysis should not be viewed as an automatic output of materials for both syllabus and tests, rather to provide a framework of some kind to draw the behavioural specifications of groups of learners in terms of what they are going to use the target language for. Detailed needs analysis generate so many micro-language skills, to the extent that it is impossible to make use of the majority of them. Weir (1984) and B.J. Carroll (1983) made use of the full needs specification they had adopted from Munby (1978), but in the actual construction of their tests did not consider so many micro-skills should be included in their tests. In this respect, B.J. Carroll (1983) wrote:

> 'Literally millions of items would be needed to tap adequately the range of linguistic communicative features emerging from the analysis. What we had to carry out was a ruthless pruning exercise so that the bare bones of the specification emerged'.
>
> (p.97)

Suffice all this is an inclusion of sub tests based on work sample analysis of the kind of work the students are required to perform at the macro level rather than the micro level.

2.14 Aims of the proposed course

The English Language Course is intended for Saudi flying-cadets of the K.F.A.A. These cadets come to the Academy, after completing secondary school, to train as pilots. With the exception of the Humanities, all training in the Academy is English medium and lasts for a minimum of three years. Training is given by native-speaking instructors with the help of a number of Saudies.

The Academy does not place an emphasis on the level of English language proficiency which an applicant should possess upon entry, but rather the overall performance in the national secondary school examination of which English is one of the subjects, in addition to other criteria such as medical fitness. An applicant is also assessed in terms of his 'mechanical aptitude'; a test is given to each candidate but this test is still in the process of validation (i.e. to my knowledge, no candidate has been rejected on the basis of his performance on this test).

The Academy has to provide the English language training for a maximum period of one year, since a large number of candidates (90%) need to be trained in the language beore embarking on their studies.

The question to be discussed is:

What could be done for these candidates who wish to undertake the Academy's course (Aerosciences and Flying Training) taught in the medium of English, who have not reached the level of English proficiency necessary for them to cope with their specialised textbooks and ability to understand and communicate with their aerosciences and flying instructors?

Briefly, the answer is in order to promote their English language

proficiency, they need to be trained in the English language. A more important issue is what should be the characteristics of the language programme? This necessarily involves a consideration of a number of factors: sociological factors which involves the characteristics of the learners relating to age, previous experience in the target language, and the purposes of what they are going to learn the language for. A second factor is linguistic, this involves the process of selection for the linguistic content. This can be arrived at by an appropriate description of the language of that which the learner is asked to perform, not only in terms of the code features of the language system but the communicative acts of language use. A third factor is pedagogic, in terms of what language skills to be taught and in what order. A fourth factor is psychological. This involves the choice of an appropriate language learning theory or theories and this can be arrived at by developing an understanding of the nature of language and language learning (for a full discussion, see MacKay (1978) in MacKay and Mountford (1978)).

2.15 <u>The learners defined</u>

The cadets are secondary school leavers aged between eighteen and twenty-three. At the secondary school, students have to choose between either the scientific department in which they have to study subjects such as Maths, Physics, Biology, etc. or the Literary department where they have to study History, Geography, Sociology, Psychology, etc. Candidates for the Academy are a mixture of the two types. All teaching in the subjects cited above is Arabic medium i.e. at secondary school level.

English is taught as an academic discipline to these students in some form or another for six years. Influenced by poor teaching and inaccurate assessment in the language, a large number of students are admitted to the Academy with a low level of proficiency.

2.16 The role of English in the Academy

In order for cadets at the Academy to comprehend and manipulate scientific materials, a high standard of English proficiency is needed. Success in their study is closely related to attaining a good proficiency in the language. In this way the role of the English language is seen to be a means in itself, it plays an auxiliary but a very important role. The cadets have to attend lectures, they need to take notes, ask questions for clarification, read their textbooks and take exams. All these activities can only be adequately performed when an individual possesses an adequate level of English language competence.

2.17 Social roles

The learners will use English as cadets in interaction with Aeroscience instructors, flying instructors, air traffic controllers and less often with fellow cadets. Although using formal register in the presentation of lessons, instructors often will have to receive and understand less formal register. This is particularly noticeable when the subject-matter dealt with is at a more personal level.

2.18 Psychological roles

Van Ek (1975) identified four types of psychological roles in the process of language learning. These were neutrality, equality, sympathy and antipathy. Learners at the Academy will normally

interact in a neutral sub-ordinate role. Being in a military Academy, in the first place, they will have to receive commands and instructions. On the other hand, they will also have to receive questions and re-state information in the context of subject-matter learning, in classroom situations, cockpits and simulators.

2.19 Settings

We are able to divide the settings of language use into:

(1) Aeroscience classrooms at the ground school.

(2) Aircraft at the flying school

i simulators

ii cockpits

With the exception of the interaction in aeroscience classroom presentation, the interaction between flying instructors and trainee cadets, trainee cadets and air traffic controllers will take place against some noise.

2.20 Topics

All topics of weight, measurement, flying instruments, engines, heat, speed, airfields, rules of the air, gravity, theory of flight, aircraft systems, chartwork, weather conditions, weather maps, earth and the language needed to describe shape, temperature, speed, will be invaluable to cadets. In addition cadets will need to identify themselves, to give and request information. They will also need to obey instructions and commands and to describe or ask about flying procedures, dimensions and positions. Cadets will also be exposed to topics which are of general interest to them.

2.21 Language skills

The following table illustrates the environment use of the different language skills.

skill	Aeroscience classroom	cockpits
spoken receptive	+	+
spoken productive	+	+
visual productive	+	+
visual receptive	+	+
written receptive	+	-
written productive	+	-

2.21.1 English skills requirements at the K.F.A.A.

Assessing language skills needed to follow a specific communication demand of academic subjects, particularly at colleges or universities where the medium is a target language, i.e. a second language, is becoming a fundamental practice among those who are in charge of the language programmes (teachers, syllabus designers and alike).

For the purpose of assessing the language skills needed to follow the Academy's course, in this case aerosciences, in the medium of English, an Academic skills Questionnaire (see appendix) was distributed at K.F.A.A. to all available instructors of English, aerosciences and flying instructors in order to determine which skills (reading, writing, listening or speaking) were most essential to non-native speakers success in aerosciences and flying.

2.21.2 The study

The study of necessary English skills discussed here was undertaken at the K.F.A.A. in July 1985. For this purpose a questionnaire was developed (see appendix). This instrument was distributed among all instructors at the Academy. Since the study covers three major schools at the Academy, the Language School, the Aerosciences School and the Flying School, it was not possible to meet with all participants who completed the questionnaire. Instead the responsible heads of schools were briefed of the purpose and the content of the questionnaire and they then briefed their colleagues on the purpose of the questionnaire. However, all participants were given an extension phone number, where the writer used to be available, for any enquiry about the questionnaire, its content or how to complete it. Although all instructors were asked not to put their names down some of them voluntarily did. In completing the questionnaire, each instructor was asked to focus on his own subject, the subject he teaches, and to rank English skills essential to follow his subject in order of importance.

The excellent response to the questionnaire (nearly from all instructors available) might be attributed to general concern about English proficiency on the part of the learners in the Academy. Before dealing with quantative data, it is worth quoting some of the comments made by instructors on the academic skill questionnaire. One English instructor commented:

> 'Practically, it does appear that for success in my classes (i.e. passing E.C.L.) the more negative aspects of language teaching need to be emphasised.'

Another English Language instructor commented as follows:

> 'Generally speaking the present administration of language training in K.F.A.A. overemphasise multiple-choice examinations leading instructors to believe that "vocabulary" is all that is necessary.'

Another commented:

> 'What is wanted for my class often is not what is wanted for success in the course (ALC),'

These and similar comments from English language instructors indirectly reflect the general concern about the teaching of English in the Academy. The concern is shown through a covert criticism of the testing system which dictates what is to be taught. It was not at all unexpected that these comments were presented. However, the negative aspect of the E.C.L. has already been presented and discussed earlier in the form of the test backwash effects and there is no need to repeat it here.

The returned questionnaires were analysed by disciplines. That is, the aerosciences instructors responses were analysed separately from the flying instructors responses, hence, it was anticipated that the ranking of the skills by the two types of respondents might be different. First, results are presented and then discussed. We will begin with the flying instructors.

For flying training, listening was considered the most essential of the four skills by 100 per cent of the instructors. They all ranked the listening skill as the first important skill. There was some disagreement with regard to the speaking ability, while 65% ranked speaking as the second most important skill to success in flying training, 34% ranked it as the first most important skill.

Generally speaking, there seems to be an agreement with regard to the importance of the oral competence in so far as flying training is concerned. Reading was also considered as the third most important skill in the flying activity. 75% of the respondents ranked reading as the third important skill. Writing was considered much less important with only one percent of all the respondents ranking it first.

Aerosciences involve a number of disciplines. Each discipline may well require a different order of skills. It was not at all surprising that there is a breakdown of responses among all aerosciences instructors. However there was a sort of general agreement in the order of skills in terms of their importance. While about 70% ranked listening as the first most important skill, only 20% ranked it as the second most important. Speaking on the other hand was ranked as the second most important skill by around 50% of the respondents while the remaining 50% ranked it as the third most important. Reading was ranked as the third most important skill by about 50%. Writing was ranked as the least important skill by 55% of the respondents.

2.21.3 Type of proficiency needed

There was another breakdown of responses between those who ranked specific proficiency in English and those who ranked general proficiency in English as of first importance. 45% of flying instructors chose general proficiency in English as being more important to success in flying training, on the other hand another 45% chose proficiency specific to discipline, the remaining 10% chose both.

So far as the aerosciences' instructors were concerned, 65% chose proficiency in general English as opposed to 30% who chose proficiency specific to discipline, the remaining 5% chose both. There could be a variety of reasons for the general English preference, the most important one is the concern of the general language proficiency in terms of constructing accurate utterances, in terms of structure and pronunciation. It might be the case that students have learnt about the subject matter but they cannot express themeselves in an acceptable form of English. The second reason, that instructors may not understand the nature of ESP, they perhaps think of it as being more related to vocabulary alone. However, I myself think that the first explanation might be the most sensible, since even among applied linguists, the definition of ESP takes different aspects.

2.21.4 Conclusions and implications

It is possible to draw a number of conclusions about language needs, syllabus design and language testing with regard to this context. The survey may indicate that in order to prepare the cadets for their Academy's study, they should concentrate on those skills which are considered most important by the Academy's instructors viz. Listening, Speaking and Reading. Systematic teaching of listening and note-taking should be an integrated part of activity in all classrooms. Teaching of speaking was shown to be of most importance and so was reading.

The survey may indicate that the E.C.L. which is primarily a test of "vocabulary" and of receptive skills, is not an adequate instrument and should be either replaced or expanded.

Having established the characteristics of the learners as well as the context of situation, we will now turn to specify the rationale behind the proposed syllabus.

2.22 The proposed syllabus

Clearly, the majority of these students we identified have a long way to go before they will be able to cope with, and profit from, an English medium course of study. One possible answer to this problem would be to provide a course in 'General English'. However, it must be noted that all these students have studied English in some form or another for six years. If a general course has not, during these six years, brought them to the level of proficiency required by the Academy, we assume that it may have prepared them for a more advanced language programme. It should be noted that although these students are of low proficiency level they do have specific purposes for studying English. Such a course would aim to be culturally acceptable to operate on topics giving rise to relevant lexis, and to provide the students with those linguistic and communication skills which will enable them eventually to progress to their field of study.

The argument above takes us to the issue of English for special purposes (ESP).

2.22.1 ESP defined

The term English language for special purposes generally refers to the teaching of English for clearly defined purposes. MacKay and Mountford (1978) refers to this purpose as 'utilitarian purpose'. Another elaborated definition is given by Strevens (1980), he wrote:

'ESP entails the provision of English language instruction:

(i) devised to meet the learners' particular needs;

(ii) related in themes and topics to designated occupations or areas of study;

(iii) selective as to language content;

(iv) when indicated, restricted as to the language skills included.'

The second feature of Streven's definition of ESP, in particular, leads us to the consideration of what has become known as 'authenticity' among syllabus designers.

2.22.2 <u>The notion of authenticity in ESP</u>

Widdowson (1978) reported that there is some divergence of opinion over whether the materials writer should use authentic texts or should create his own simple texts for teaching purposes. By authentic is meant that the input materials of the language syllabus such as topics and notions are taken unchanged, or to use Hutchinson and Waters' term (1980) 'unadulterated' from the learners' scientific or otherwise course of study. ESP materials writers tend to make use of what is referred to as 'carrier content', the extraction of materials for the language programme unchanged. However, the use of authentic materials despite its apparent face validity may create some problems on the part of the learner as well as the language teacher. The learner might not need all the lexical items generated by authentic materials prior to starting his chosen subject, the danger, therefore, might be that the learner will devote most of his time to acquiring these lexical items at the expense of other skills presented in the course which he will certainly need and will employ in his course of study.

As far as the language teacher is concerned, the problem might be even worse. We do not assume that the native language teacher knows all concepts and vocabulary items of all scientific disciplines or indeed in any one discipline; the problem might be even greater when the language teacher is a non-native speaker. Anyway, my own view is that making use of the learners' chosen subject is desirable, but input materials for the language syllabus should be simplified specially if the syllabus is intended to cater for different levels of competences. The process of simplification can be carried out with the help of EFL experts and the subject-matter teacher so that the language teacher does not end up teaching the subject matter. A consideration of the materials for the language programme is only one aspect of the matter. Another aspect is the subject-matter classroom observations in order to gain insights of the 'teaching strategies' employed by the subject matter teachers as well as the 'linguistic competence' (Hutchinson and Waters, 1980) they assume the learners know. All these are very important factors in the process of ESP syllabus design. They all contribute to the language input materials, learning strategies and teaching methodology.

2.22.3 Aerosciences classroom observations

It was observed that the main form of presentation in Aerosciences classrooms, such as Aeroengine, Aerodynamics, Navigation, is audio-visual. The instructor does not just give a lecture about the 'ejection seat' for example or the 'jet engine'; he uses either the actual machine or a model as an integral part of the presentation. The use of the audio-visual as a language teaching method may well

be desirable in this context.

2.22.4 Theoretical assumptions and pedagogic implications

From the arguments outlined above, the theoretical assumptions about language learning and teaching are based on recent developments in theories of Communicative language teaching which are represented in the work of Hymes (1972); Canale and Swain (1980); Munby (1978); Widdowson (1978, 1983); Wilkins (1976); Brumfit and Johnson (1979); Yalden (1983), and so many others. These theoretical assumptions are formulated in the objectives and specifications of the proposed course.

2.22.5 Overall objectives of the proposed course

A new language course is needed to replace the existing (ALC) one. It should aim to prepare students to enter their field of study. It has to equip the students with the basic linguistic and communication skills necessary to carry out their subject matter, so that within one year students can effectively proceed with their field of study. This requires an examination of the kind of work the students are expected to do and be able to cope with, once they master the language in order to gain some insights into the type of materials they will handle together with its features as well as the teaching strategies the subject matter teachers employ, as well as what they assume the students know, in terms of the linguistic competence and knowledge of the world.

(1) The input materials should be taken from all the different subjects to which the students will have been exposed. These materials should be simplified to suit the EFL teachers and learners. It should be flexible enough to cater for the

diversity in linguistic competence which exists among students.

(2) Communication implies a knowledge of language usage as well as language use. Usage is a prerequisite for use therefore competence in both are necessary and required.

(3) Communication calls for the integration of language skills. Language skills should be taught integratively when necessary. Note-taking for example, is an activity which needs to be developed. As a study skill it involves two skills, listening and writing. In this case the development of note-taking implies the teaching of these skills in combination with one another.

(4) Demonstrations and lecture presentations in Aerosciences' classes are oral and visual related. Students need to develop a fluency in handling visual-related discourse. This entails learning how to structure information logically as well as how to make use of the visual clues of the presentation.

So far we have been talking about the general characteristics of the proposed course. Below, we are going to deal with the general objectives of each of the four skills.

I <u>Reading</u>

The goal: to read authentic, specialist textbooks (Aerosciences) in English with understanding.

<u>Objectives</u>:

(1) To use linguistic information in the written text as clues to meaning. This includes recognition and interpretation of formal cohesive devices for linking different parts of a text, deducing the meaning and use of unfamiliar lexical

items through an understanding of word formation and context clues, and decoding of syntactic structures.

(2) Reading for information.

(i) identification of topic, main ideas and reading for detail.

(ii) skimming to obtain the gist of the semantic content

(iii) scanning to locate specific information.

II Listening/speaking

The goal: enabling students to understand and communicate orally in English with their flying and aerosciences instructors.

Objectives

(1) Ability to compose and understand comprehensible utterances by manipulating such linguistic micro skills as subject-verb concord, tense harmony, post modification, clause linkage.

(2) Ability to interpret the intended message through an understanding of the basic intonation features.

(3) Articulation of isolated sounds and connected speech.

(4) Expression of explicit information and conceptual meaning.

(5) Describing objects and processes.

(6) Reporting information.

(7) Answering and asking questions.

III Writing

(1) Manipulation of script.

(2) Extraction of information from text materials.

(3) Conversion of charts and diagrams to writing.

2.22.6 Summary and Conclusions

In this chapter we have attempted to argue that the motivation of the learners is one of the most important factors affecting success or failure in language learning. In our discussion we identified three stages where the presence of motivation is necessary and required. With regard to our situation at K.F.A.A. we spoke of a pre-language course motivation, an in-course motivation and a post-course motivation. We concluded that upon entry, our cadets are highly motivated in the sense that they realise they are in need of training in the English language and that their background in the target language does not enable them to proceed to an English medium course just yet.

In our distinction between integrative and instrumental motivations we argued that our learners are not integratively motivated by aspects of the American life styles (to which the present course, the ALC, makes reference) or even to topics which are not culturally oriented and of no potential value for the students such as "lumber, supermarkets, tableware, at a drug store"... etc. which are bound to generate so many irrelevant lexical items to the learners. Our learners would like to get at their chosen field of study through a knowledge of the target language which they see as relevant to what they are going to do with the language.

We also outlined some additional shortcomings of the ALC. We found out how the course is heavily weighted towards the acquisition of irrelevant lexical items. Specific objectives of each unit seem to emphasise just that. As a first objective we also identified the

ALC as a linear type of programme. We then argued that a cyclic approach is more preferred to a linear approach. A cyclic oriented course would cater for the diversity in students progress, since the rate of progress among individual learners varies.

We then moved to discuss the construct theory of the ALC. As a structural oriented syllabus, the ALC focuses on teaching the language as a stock of 'usage'. Recently, language teaching has shifted focus from teaching the language as 'usage' to teaching language as 'use', from an atomistic approach to a sociolinguistic approach. This change has implied a focus of teaching the foreign language as communication rather than teaching the grammatical system of the language.

However, teaching the language for effective communication should in fact emphasise both accuracy and fluency.

What we have been witnessing, in the field of language teaching is that people (syllabus designers, teachers and language policy makers) tend to go with the 'fashion' i.e. when a new theory in language learning/teaching has appeared, people do not start speculating about it, rather they adopt it and later when they discover its deficiencies, they start criticising it. The history of language teaching is full of examples.

Out of our arguments, the construct theory we proposed is inferred from the following: Since the ultimate aim of language learning is communication, effective communication cannot be attained without an underlying knowledge of the system of the language, this knowledge can be achieved by fulfilling the immediate objectives of the language teaching situations. Immediate

objectives can be identified through the students' immediate language problems which might be weaknesses in the knowledge of part or whole of the system of the language which they may need to acquire first, in this sense teaching language as Usage may serve as a prerequisite for effective communication.

We then proposed a set of specifications for the language syllabus our learners at the K.F.A.A. need. We considered in some detail a number of factors involved in the design of the language programme. These were sociological factors which involved the characteristics of the learners, in terms of age, the background knowledge of the target language and the purposes and the uses of the target language within the Academy's context of situations. We also considered the linguistic content and the pedagogic factors in terms of the language skills needed by the students. We called for an integration of language skills since a natural communicative event involves employing more than one skill.

We argued for a more E.S.P. oriented type of course rather than a general language course. A course which purports to be culturally acceptable and relevant in content to the actual needs of our students. It was suggested that input materials for the language course should be drawn from the students' field of study, from all the Aerosciences subjects they are going to deal with. A number of topics for the language programme was suggested. We also suggested that input materials should undergo a process of simplification, the reason being, we argued, that a learner may not need all the technical vocabulary items prior to starting his specialised training. At the same time, we appealed for face and content

validity which the present ALC does not take account of. We also added another dimension to the present construct theory and that is teaching the 'Use' of the language in addition to 'Usage'.

CHAPTER THREE

THE STATE OF THE ART OF LANGUAGE TESTING

REVIEW OF THE LITERATURE

3. The testing of second language proficiency tends to follow developments in linguistic theory, as well as language teaching methodologies (Davies 1968, 1977). This chapter attempts to report on the advances made in the field of language testing. As a point of departure, the classification, made by Spolsky (1975, 1978) of the three distinct but overlapping trends in the field of language testing is discussed and exemplified, the trends to be discussed are: the pre-scientific, the psychometric-structuralist and the sociolinguistic-psycholinguistic trend. The work of some, well known, language testers is cited and reviewed. The work of Lado (1961), Davies (1965), Ingram (1963) and B.J. Carroll (1978, 1980) is presented and discussed. Analysis of some standardized tests which represent the influence of the second and third eras is also provided. The proficiency tests reviewed are the T.O.E.F.L., the E.P.T.B., the E.L.B.A., and the E.L.T.S. The theoretical principles underlying the construction of these tests are outlined and discussed. Dictation and cloze receive some discussion. Testing for specific purposes receives some discussion too and measuring speaking ability is considered in some detail. The concept of validity is reviewed and some validation studies made on proficiency tests are provided and discussed. The notion of proficiency as well as the traditional and current aspects of assessing proficiency in language is considered in some detail.

The final section in this chapter is a short summary of the

state of the art.

3.1 Introduction

When designing or even selecting a test of foreign language proficiency one has to consider the purpose for which the test is to be used. No single test can be claimed to possess universality.

Analysis of the nature of the objectives and use on which students will be assessed is necessarily the first step before setting about to construct or even to choose an existing proficiency test. The result of the analysis can then have a direct and an important influence on the linguistic content of the test being developed as well as the skills or tasks to be tested. An important decision also has to be made on the trend of proficiency to be adopted and on the basic design of the test.

The central issue in assessing proficiency is determining what is meant by proficiency? Moller (1981) commented 'there is no clear answer as yet'.

Over the last two and a half decades several approaches to language proficiency have been identified. These approaches were inaugurated as a result of developments in linguistic theory and language testing research.

The first trend, the second of Spolsky's classification, derives from structural linguistics and "contrastive analysis" (Lado 1961). Proficiency within this approach is viewed as:

(1) The ability to manipulate features of the systems of the language concerned i.e. phonology, grammar, lexicon and demonstrating the ability to operate in some or all of the skill areas i.e. listening, speaking, reading and writing.

The notion which underlies this approach is that demonstrated ability to recognize a large number of discrete item tasks will add up to a suitable performance when the different components of the language and skills are put together.

Another approach views proficiency as an overall. Oller (1979) referred to it as "the grammar of expectancy approach". The testee is called upon to demonstrate his abilities to operate simultaneously the grammar of the language, i.e. different levels of the language, phonology, syntax, morphology and the lexicon. This approach sees language as a means of communication. What is assessed is the student's ability to use the language.

(All these approaches together with the pre-scientific trend are discussed thoroughly below.)

Techniques of assessment in these approaches are as follows:

(a) for the psychometric-structuralist approach, discrete point tasks are presented to the testees. Examples of these tests are the E.P.T.B., the T.O.E.F.L. and the E.L.B.A., or what is called the standardized test batteries.

(b) Overall proficiency is tested through cloze and dictation.

(c) Communicative competence is tested through structured interviews, role plays and other specific oral tasks. An example of communicative tests is the E.L.T.S.

(Below is a description of the linguistic content of some of these tests.)

3.2 Approaches to Language Testing

In the light of developments in linguistic theory and language teaching and testing research, Spolsky (1975) identified three

phases in the history of language testing: (1) The pre-scientific, (2) The psychometric-structuralist, and (3) The psycholinguistic-sociolinguistic approach. Moller (1981) adds a fourth phase (4) Communicative. (Wilkins 1976), (B.J. Carroll 1978, 1980), Munby 1978).

3.2.1 The pre-scientific (traditional tests)

(a) Forms of these tests - traditional tests take the following forms:

(i) Translation from the native language of the learner to the target language, and from the target language to the native language. Long texts are normally given, but single sentences are also given to translate from the native language to the target language and vice versa.

(ii) Sentence completion.

(iii) Dictation.

(iv) Oral interview.

(v) Compositions in the target language at an advanced level.

3.2.1.1 (b) The characteristics of these tests

(i) Assessment in all these types of tests is highly subjective.

(ii) The objectives of tests are not pre-determined.

(iii) The proficiency being tested is not clearly identified (Moller 1981).

(iv) The language teacher is the one who is involved in preparing these tests as well as marking them.

In brief, the pre-scientific approach to language testing does not take into account the basic concepts of a good test. It lacks the two main features which a good test has to possess (1) reliability

and (2) validity.

3.3 The psychometric structuralist approach involved the entry of the experts into the language testing field. This approach derives from structural linguistics and the contrastive analysis technique.

Learners are presented with individual features of the various systems of the language i.e. syntactic, phonological and lexical items. The learner's role, then, is to manipulate these features of the system of the language being tested. The notion behind this approach is that students are presented with a large number of discrete point tasks to perform. These tasks will add up to adequate performance when the features are put together in connected speech i.e. discourse. Items are presented at the sentence level or even lower.

Proficiency within this framework is viewed as 'demonstrating the ability to operate in some or all of the skill areas of listening, reading, speaking and writing' (Moller 1981).

Objective tests were developed, the most widely used being the multiple-choice item, and analytic scoring procedures were introduced. Statistical methods were also introduced and improved to measure the reliability and validity of tests.

Before giving a description of some of the tests constructed at this stage, let us first consider Lado's great contribution to the field of language testing.

3.3.1. Lado

No-one denies Lado's great contribution to the field of language testing represented by his book 'Language Testing' (1961). Lado's work has been acknowledged by writers in the field of language

testing, such as Spolsky (1978) and Heaton (1975). Spolsky (1978) wrote:

> "It is not too much of exaggeration to suggest that a great proportion of work in language testing since Lado's book appeared is either based on it or an attempt to answer or correct some of the points he made."
>
> (Spolsky 1978, p.220)

Heaton has implied that it is impossible to work in the field of language testing without referring to Lado (1961).

One main feature of Lado's approach is the separation of the complexities of language into segments. This approach affects what is to be tested and how this testing is to be carried out. A structural contrastive analysis between the learners mother tongue and the target language serves as a filter as to what is to be tested. According to Lado, a structure is not to be seen as that of grammatical structure (syntax), but it is the different components or levels of the language as a whole. These levels are morphology, syntax, lexicon and phonology.

Through contrastive analysis one could determine:

(1) What language patterns are similar in both L1 and L2.

(2) What patterns in the target language have no equivalent forms in the learner's native language, which may pose a problem in use.

The resulting are discrete item tests which take account of the contrastive analysis process. Each item handles one level of the language in terms of one of the four skills. According to Morrow (1981) 'one main advantage of this form of test construction is that it yields data which are easily quantifiable'.

As far as the psychological implications are concerned, Lado's tests raise questions as to whether or not correct habits have been established. According to Lado passing a test should involve a specified proportion of accurate responses. In other words, Lado's approach favours the responses to be assessed quantitively rather than qualitatively.

Although Davies (1978) remarks that 'there is very often a gap between what Lado himself does and what he says he does' his work remains a major reference in language testing.

3.3.2 Assumptions underlying the structural approach

The assumptions, therefore, which underlie the structural psychometric approach to test construction depends on the view which sees the knowledge of the elements of the language as equivalent to knowledge of the language. One main criticism of this assumption is that to know the elements of the language does not mean a lot unless the learner has the command to put these elements together to produce appropriate utterances in accordance with the situation in which the language will be used. Knowledge of the elements of language alone is, in fact, not enough unless the testee is able to combine them in appropriate linguistic contexts in which he will use the language.

Despite this criticism parts of most of the standardized language tests of proficiency have been and are still being constructed in this manner i.e. the complexities of language are analysed into levels and skills and are tested independently of each other. Examples of these tests are the 'T.O.E.F.L.' developed by the Educational Testing Service, the 'E.P.T.B.' devised by Alan

Davies (1964, 1965), 'Michigan Test' devised by Upshur and John (1961) and the 'E.L.B.A.' constructed by E. Ingram and used by the University of Edinburgh.

More recently a group of test-constructors working under the Education Ministry of the People's Republic of China devised a new Test Battery which they called English Proficiency Test (E.P.T.). It has been in operation since 1980. It is a standardized test. The test is being mainly used to select scholars who aim to study overseas. In addition to the traditional components of standardized test batteries i.e. grammatical structure test, vocabulary, reading and listening tests (mainly discrete item tests) cloze and guided writing have been incorporated into the test.

The proponents of the psychometric view have worked with the traditionalists of the prescientific trend to develop more reliable and objective means of scoring writing tests and oral tests (Valette 1977). The problem with these tests is one of economy. The scoring of composition tests (essay type of tests) and the administration of oral tests require well trained scorers as well as a considerable amount of time.

Through the following statement Lado stresses the importance of contrastive linguistic analysis to determine the content of the test. Moller (1982) quotes Lado (1957):

> 'The advance in English language testing came not from connected material but from concentrating on the language problems as they actually are. And we get closest to the language problems by a systematic comparison on the native language and the foreign language'.
> (Lado 1957:4)

While Lado recognises the importance of the qualities of good

tests i.e. validity and reliability, which were introduced by psychologists, he insisted that what determines the content of tests is linguistics and not statistical treatment. He viewed the statistical analysis as a contribution to the 'refinement of the test' (Lado, 1957).

A typical English proficiency test battery designed in accordance with the psychometric structuralist approach would include some or all of these sections.

1. Listening comprehension

   stimulus: (spoken): sentence, conversation or monologue

   response: selections of an appropriate answer from among 3 to 4 alternatives written in the test booklet

2. Vocabulary (some test batteries do not include a section on vocabulary, perhaps there was no adequate list of vocabulary sampling e.g. E.P.T.B. (1964, 65). If there is a section, then, it will be as follows:

   stimulus: (written): single word in context of a sentence or single word.

   response: selection from a number of alternatives of either the most appropriate meaning covered by the word in question or a synonym.

3. Reading comprehension

   stimulus: (written): paragraphs of connected prose followed by questions.

   response: selection of alternatives of an appropriate response.

4. Structure (grammar)

   stimulus: (written): one or two sentences, often in context of a conversation.

   response: selection or recognition of an appropriate word or phrase from among a number of distractors.

The scoring of these kinds of tests is completely objective, the

total number of items would be between 150-200 and the whole test would take 1 to 2½ hours.

3.4 Descriptions of some test batteries

3.4.1 English proficiency Test Battery (E.P.T.B.)

Davies (1965) reported that the British Council was dissatisfied with its subjective method of assessment of English proficiency. Adding to that the frequent complaints about non-native students of English learning in the United Kingdom whose inadequate English has frustrated their courses of study. These were among the factors which led to the thinking of designing a new test battery that would serve as a filter to assess the learner's 'real' language proficiency before admitting him to his course of study. Davies (1965) reported that the conclusions of his investigation indicated a high degree of unreliability between the assessment of students' English on the British Council subjective method of assessment and his 'real' proficiency when faced with a full workload in the British Universities.

All of the above factors reinforced the urgent need for constructing a new test that would, to put it in Davies' terms, 'assess all the important aspects of the use of English' under the outlined conditions.

The construction of the original (E.P.T.B.) was designed on the psychometric trend of work sampling and on the basis of certain linguistic skills and components to be tested.

The impression I had from Davies' (1965) statement '... to assess all the important aspects of the use of English under conditions' is that he was aware of the fact that productive skills

should be tested, but the practical difficulties of administering these kinds of tests together with the realiability factor have prevented him from doing so.

3.4.1.1 Overall structure and components of E.P.T.B.

Test 1 listening, phonemic discrimination (vowels and consonants) words in isolation

Test 2 phonemic discrimination - words in sentences

Test 3 listening, intonation and stress

Test 4 listening comprehension

Test 5 reading speed reading

Test 6 reading comprehension

Test 7 reading grammar-recognition of appropriate features

All the above subtests contained discrete items which could be scored objectively. The highest possible reliability was aimed at for all the subtests. (Only tests 1, 3, 6, 7, the 'short version' were used by the British Council.)

Below we will examine the theoretical concepts underlying the construction of the E.P.T.B. in some detail.

3.4.1.2 The English Proficiency Test Battery 'E.P.T.B.'

The E.P.T.B. was designed in the early '60s. The design of the test Battery was based on two assumptions. One linguistic, the other is psychometric. The linguistic approach derives its input from the flourishing, then, structural approach to language testing/ teaching. The psychometric assumption is with predictive validity.

Although it was recognized from the outset that tests of oral and written production are of the highest importance to students coming to study in this country (Britain), these tests were not

included in the Battery due to the practical problems involved in the administration of these tests and problems due to reliability of scoring procedures.

The E.P.T.B. was originally designed to measure the reading and listening comprehension skills of students coming to study in this country. [For a complete description of the theory underlying the construction of the E.P.T.B. see Davies 1965.]

The original Battery is composed of the following sub-tests:

1.	Listening	phonemic discrimination words in isolation
2.	Listening	phonemic discrimination words in sentences
3.	Listening	intonation and stress discrimination
4.	Listening	comprehension work sample
5.	Reading	speed reading
6.	Reading comprehension	
7.	Grammar	recognition of appropriate usage

3.4.1.3 Format

All sub-tests in the E.P.T.B. contained discrete item tests and were scored objectively.

3.4.1.4 Validity

Concurrent and predictive

To my knowledge, prior to its release for operational purposes, E.P.T.B. was the only test among the well-known standardized tests which provided a lot of normative data. Reliability, concurrent and predictive validity were claimed for the E.P.T.B. These studies were made known to test consumers prior to its release for operational purposes. However, the test is not consistent with

current thinking of linguistics language teaching and testing. At the same time it is maintained that current thinking i.e. tests marking the psycholinguistic-sociolinguistic era may not be better off or at least there is no evidence to substantiate this as yet.

The evidence obtained so far is that these tests may be as valid concurrently as linguistic based tests. Moreover they emphasise the use of language. The validation studies reported so far suggest that the E.P.T.B. overlaps with E.L.T.S. A correlation coefficient value of .84 was noted between the E.L.T.S. and E.P.T.B., which may suggest that the two tests rank students similarly.

3.5 <u>E.L.B.A.</u>

E.L.B.A. is a proficiency test of English as a foreign language. It was designed to distinguish between such students who are able, from a language view point, to follow a specialised course at an institute of higher education, where the medium of instruction is English and those who are not. It is an objective multiple choice test.

E.L.B.A. consists of two parts I and II. Part I is composed of 150 items divided into four subtests.

(a) Sound recognition (100)

(b) Intonation (10)

(c) Stress (10)

(d) Listening Comprehension (30)

Part I

(i) Listening:

Phoneme discrimination (vowels and consonants). A word is only

read once and the testee chooses from among three distractors, e.g.

1. heat hit hat

2. phase fez fares

(ii) Intonation:

George has just left. The three choices are

1. a straightforward statement

2. asking a question

3. very surprised

The student is asked to identify the type of intonation.

(iii) Stress:

What is your name?

( ) ( ) ( )

The student is required to label the syllable that carries the heaviest tonic.

Listening, Multiple choice

A question or a first part of a conversation or an incomplete sentence. The testee is required to choose the right answer.

Part II: (has three subtests)

1. Grammar (50 items)

2. Vocabulary (50 items)

3. Reading Comprehension (20 items)

e.g.

It is hot in India,

(a) does not it

(b) does not he

(c) Isn't it

(d) it is

2. Vocabulary: Deteriorate (words in isolation)

choose the synonym (a) prevent (b) become worse

(c) criticize (d) slow up

3. Reading Comprehension:

passages followed by several questions about each passage.

3.5.1 Reliability

High reliability in all subtests were aimed at. Reliability (Kuder Richardson) N = 293 as can be seen from this table: are, in general, very satisfactory.

Sound discrimination	.87
intonation	.62
stress	.70
listening comprehension	.85
Part I listening	.93
Grammar	.93
Vocabulary	.93
Reading Comprehension	.81
Part II reading	.96
Test total	.97

Predictive validity of E.L.B.A. will be discussed in the validation studies section in this chapter.

3.6 T.O.E.F.L (U.S.A.)

The experience of E.F.L. testing in the U.S.A. paralleled that in the U.K. By 1960 there was an urgent need for a good method of assessing the English language proficiency of foreign students applying for universities in the U.S.A. Spolsky (1979) reported

that a survey which was conducted in 1962 among Colleges and Universities revealed that more than 90% of the respondents felt the need for a better programme of testing the English of foreign students.

In 1963, test planning and test writing committees met, established details of the test, and prepared to begin the writing of test items. The test was labelled T.O.E.F.L. The test takes a multiple choice format.

3.6.1 Theory underlying T.O.E.F.L.

Three principles were agreed upon by the planning committee of the National Council of Teachers of English as a foreign language (N.C.T.E.). These principles are:

1. Contrastive analysis as a basis for test construction was rejected. This technique as we saw earlier was proposed by Lado (1961). The rejection was apparently because of the heterogeneity of learners as well as the impracticality of accounting for all the hundreds of language backgrounds of the learners.

2. Vocabulary control was to be exercised whenever possible in order to keep reading comprehension at a low level of difficulty.

3. The test items were to be written in a way that would present language 'in as realistic a context as possible' (N.C.T.E. 1963:2).

A typical T.O.E.F.L. test, in its original form, consists of five sections. Each section is composed of subparts.

Section I

Listening Comprehension

Part A requires the testee to recognise the most appropriate

answer among four distractors. A testee hears either a statement or a question, and he is to be able to demonstrate his ability to understand spoken English.

e.g.

When did Tom come here?

(a) by taxi (b) yes, he did

(c) to study history (d) last night

Part E: A testee hears a conversation between two speakers. At the end a third voice will ask a question about what has been said.

e.g.

Hello Mary. This is Mr. Smith at the office.
Is Bill feeling any better, today?

Oh, yes Mr. Smith. He's feeling much better.
But the doctor says he'll have to stay in bed
until Monday.

Q. Where is Bill now?

The testee will read:

(a) At the office (b) On his way to work

(c) Home in bed (d) Away on vacation

Part C

The testee hears a lecture about a topic. While a testee listens he should take notes. At the end he is required to answer a series of questions.

Section II

English Structures

This part is designed to measure control of the grammatical structure of English. The testee is presented with a short written conversation betweem two speakers, part of which has been omitted.

Four choices are written beneath the conversation. The testee is required to choose the appropriate answer.

e.g.

"The coffee is very cold"

It needs

(a) hot (b) to heating

(c) to be heated (d) to be hotted

Section III

A sentence in which one word is omitted

e.g.

Jack did not have to wind the clock this morning

because he had done it

(a) previously (b) heartily

(c) consequently (d) merrily

Section IV

Reading Comprehension

A series of paragraphs, each paragraph being followed by several questions.

Section V

Writing abilities

Part A. The testee has to be able to recognise language which is inappropriate for formal written English. The testee is presented with the following:

e.g.

The student knew that his teacher expected a exact answer, but he could only reply with a guess.

Part B. A test of style. The testee is asked to complete the following:

e.g.

The young boy was pushed by forces

(a) beyond his control (b) over his own control

(c) outside of his doing it (d) without his control on them

It is not quite clear how these items would test writing ability. These items do not require students to perform tasks which we identify as 'writing' which involves generating ideas, elaborating, describing and organizing these ideas through employing a variety of syntax and lexical items. It is interesting that T.O.E.F.L. will include an essay test of students' academic writing proficiency. This sub-test as was reported by Greenberg (1986) will be introduced as an additional part in this session 1986-1987. It aims to assess students' communicative competence in writing. It is scored holistically by an overall impressionistic six point scale. The main assumption underlying its construct is that "writing competence is a situational construct which varies according to the particular situation" (Greenberg 1986). What seems will be offered are a series of field-specific tasks.

3.6.2 The new version of T.O.E.F.L.

In 1973, Pike completed a research project which aimed at, to put it in his own words, 'obtaining information useful for evaluating and possibly revising the specifications and content of T.O.E.F.L.'.

This study investigated three questions. These are

(1) How valid is the current form?

(2) What should T.O.E.F.L. measure? and

(3) What subtests should T.O.E.F.L. contain?

Angelis (1979) reported that Pike used 4 experimental multiple choice tests, 2 objective open-ended tests, and 2 subjective evaluation measures. One of the experimental tests is taken from the listening comprehension section of T.O.E.F.L. (original form, see descriptions of T.O.E.F.L. above).

The second experimental subtest required the testee to substitute an underlined word or phrase with one of four distractors. The third experimental test took the form of embedded sentences which required the testee to select the most appropriate paraphrase of several short simple sentences. The fourth test is a modified cloze procedure. Words were deleted but 4 distractors were provided for the deleted words of which one is the original word. The 2 open-ended formats were composed of a conventional open-ended cloze task and a sentence combining a task similar to the type reported by Hunt (1970). The student reads short simple sentences and is asked to rewrite them in a more acceptable format using a more complicated structure. In addition to all these tests, Pike conducted two experimental subjective measures. Testees took an oral interview test, and each was asked to write four short compositions on four different topics. The former was evaluated by 10 listeners, and the latter by 2 readers.

Pike, then, concluded that T.O.E.F.L. could be shortened and 1 or more sections could also be omitted. A high correlation between the English structure and writing ability subtests was observed, and because of their similarities in other aspects, Pike recommended

that one of them should be eliminated. Pike recommended that the listening comprehension section be retained in its original format but he advised that the lecture section be substituted for shorter passages or mini-lectures to parallel the Reading Comprehension subtest. Thus, the new T.O.E.F.L. could become a 3-part rather than a 5-part test.

The planning committee has approved of Pikes' recommendation and the T.O.E.F.L. The T.O.E.F.L. now appears in three sections rather than five.

3.6.3 Reliability of T.O.E.F.L.

A further study of the reliability of the test in its new format was conducted. Angelis (1979) reported that data collected between 1976-1977 on the new format showed that the statistical nature of the test is consistent with the previous version. The reliability figures for the 3 parts of the test range from .86 to .91, with .95 being the reliability for the total test. The Standard Error of Measurement for the 3-parts ranges from 2.46 to 3.21, with the figure for the whole test being 16.06. Inter-correlations among the 3-parts range from a low of .65 between listening and structure and written expression to a high of .81 between structure and written expressions and reading and vocabulary.

3.7 Characteristics of psychometric-structural tests:

(1) Focus on discrete linguistic items, i.e. phonology, grammar, lexicon, and integrated skills, listening comprehension, reading comprehension and writing.

(2) Emphasis on the concepts of test reliability and validity.

(3) Use of objective tests and objective scoring.

(4) Bias towards testing the receptive skills, i.e. listening and reading and testing the linguistic components through receptive skill tasks.

(5) Linguists and psychometricians play an important role in determining the content of the test.

3.8 The psycholinguistic-sociolinguistic trend focuses on 'integrative ' tests. Valette (1977) and Davies (1978) called them 'global' tests. This approach to language testing takes into account the total communicative effect of the message. The proponents of this view particularly psychologists insist that there is such a thing as overall language proficiency. Sociolinguists on the other hand emphasise the communicative competence and introduce the concept of a situational approach to testing.

The tester, then, will want to take into account elements of both the psychometric and the psycholinguistic-sociolinguistic views. The tester will want to test objectively whether his students have mastered the several components of the system of the language being learned, at the same time observe to what extent they are able to put together these components in concrete situations for the purpose of communication. Davies (1978) puts it 'the third of Spolsky's stages, the psycholinguistic-sociolinguistic emphasis is on the integrative or global test which attempts to assess proficiency in both (production and in comprehension) of the total communicative effect of the message'.

The shift from the psychometric-structuralist approach to the psycholinguistic-sociolinguistic is as a result of a serious concern as to whether the standardized test batteries (E.P.T.B., E.L.B.A.,

T.O.E.F.L., etc.) actually measure proficiency. It has been argued that students' ability to engage in satisfactory communication, either spoken or written is not being evaluated i.e. assessed. Underlying this claim is the fact that most existing standardized tests assessing proficiency are inappropriate for the purposes for which they are being used. Moller (1982) puts it as 'there is a concern regarding the validity with its four kinds'. The general argument for the preference of integrative testing is that language as used in communication forms a whole. When language is used in concrete situations, different components of language are integrated and used together. As a result language must be 'integrated' (Davies 1978) and tested simultaneously.

Technique of assessment in this approach is through cloze and dictation. This phase in language testing can be described as having the following features:

(a) Stresses the notion of overall proficiency

(b) Stresses the functional and communicative use of language

(c) Stresses the concepts of validity and reliability.

Since research is still being undertaken, there is no typical test format as yet.

3.9 Cloze as an integrative test

Alderson (1978) in a thorough survey of the literature was able to trace the idea of gap filling to the year 1897, long before Taylor (1953) introduced the cloze as a measure of readability of texts. Alderson referred to this period as Pre-Taylor.

Alderson reported that Ebbinghaus (1897) was interested in this technique (which he named "Knombinationsmethode"). He used this

technique for the measurement of intelligence. Alderson has also reported that such technique was further studied by researchers such as Brown (1910), Ballard (1920), and Hamilton (1929). Their main interest was the usage of the sentence completion technique as a measure of verbal intelligence rather than a measure of either readability of texts or of reading comprehension.

Taylor (1953) was accredited for being the inventor of the standard cloze procedure technique. He first introduced it as a device for measuring readability of texts, then, he (1959) suggested that cloze can also be used as a measure of reading comprehension with foreign- language learners. Anderson (1971) used cloze as a criterion reference test in an attempt to determine the readability of texts for English as a foreign language. Jongsma (1971) claimed that a cloze is a valid test of reading comprehension for foreign language students. Studies with non-native speakers suggest that cloze correlates well with measures of English as a foreign language proficiency (see Oller and Conrad (1971), Oller (1973), Oller, Atai and Irvine (1974), Stubbs and Tucker (1974), Aitken (1977), Streiff (1978)). The work on cloze as a measure of English as a foreign language, as reported in these studies indicated validity coefficients ranging between .71 and .89.

Oller (1972) claimed that cloze relates more to integrative type tests dictation and reading comprehension tests than to traditional, as he called them "discrete point tests" of grammar and vocabulary. He also claimed that cloze is very useful for the measuring of global skills. Oller's conclusions have led some language testers to regard cloze as an automatically valid technique of overall

language proficiency. More recently cloze has been used as a teaching device.

Generally speaking, cloze test has been accepted as an integrative test of reading comprehension (Davies 1978). Moreover, the proponents of the cloze test expressed their desire to find a single integrative measure of an overall proficiency. Davies (1978) commented that "such a desire may, of course, be vain". In other words, although the cloze has been proposed to test varieties of tasks, it has not been proved to qualify as a single test which would determine an overall English proficiency. Recent findings (Alderson 1978) showed that individual cloze tests vary greatly as measures of English as a foreign language proficiency. Alderson's (1978) findings showed that cloze in general relates more to tests of grammar and vocabulary [low order skills] (E.L.B.A. tests 5 and 6) than to tests of reading comprehension (E.L.B.A. test 7). He concludes that cloze is more related to core proficiency.

There are a number of issues in the literature of the cloze procedures, these are:

(1) The selection of texts

(2) Deletion rate and the type of deletion

(3) Scoring procedures, and

(4) Validity and reliability

We will examine these issues in turn:

3.9.1 (1) The selection of texts

Any text cannot just qualify as a good cloze test, this may suggest that cloze texts need to be validated against external criteria. A cloze test constructor has to consider several factors

before choosing a text, for testing purposes, for a particular group of learners. Among these considerations and constraints are:

(1) The intellectual content of the text

(2) The cultural content

(3) The linguistic difficulty

(4) Register and level of formality

(5) Idiosyncracies of style e.g. lists of items, idioms, proper names, dates, numbers

(6) Length of the passage

The question, therefore, is one of content validity, Davies (1978). As far as whether the texts should be known or unknown would have an effect on the mean scores, remains a controversial issue. Oller (1972) and Davies (1978) quote Jongsma (1971): 'It is known that if a student is familiar with the text of a cloze passage he will perform significantly better, and in proportion to his familiarity with the passage'. A recent study conducted by Alderson and Urquhurt (1983) confirmed Jongsma's claim. Davies (1978) reports: Hafner (1964) suggests the differences between mean scores of texts which had been taught previously and those which had been read previously was not significant and that known and unknown texts correlated highly with one another.

3.9.2 Deletion rate

The most commonly used deletion type is every nth word where nth as found in the literature varies between every 2nd to every 24th word deleted, but perhaps the most commonly used deletion rate is every 5th word. Although there has been little research on the effect of deletion rates on the mean scores, studies undertaken seem

to be contradictory. While Oller (1972) found that deletions between every fifth and every twelfth word keep results stable, Alderson (1978) on the other hand claims that deletion rates do change the results and that the cause for this change reflects the change of text with the change of deletion rates rather than change in the amount of text.

Several modified types of deletion have been proposed. These modifications fall along the continuum from productive to receptive testing, multiple choice being at the receptive end, Davies (1978). Rational deletion (every category member) has also been proposed. The criticism that these modified types has received is that these tests cannot be called cloze. Jongsma (1971) claimed that an nth rate deletion tends to correlate more highly with comprehension measures than would categorical deletion type.

3.9.3 Scoring methods

There are two well known scoring procedures, 1. the exact word replacement, the original word used by the author. This procedure has been named as the verbatim, and 2. contextually appropriate replacement. There are several kinds of contextually appropriate words, (a) grammatically appropriate, (b) semantically appropriate or (c) both together. Studies have shown that these different scoring procedures are equal in discrimination (Anderson 1972). They also correlate highly with one another. "An acceptable replacement word mean score is always being higher than the verbatim word mean" (Davies 1978). A large element of the high correlation between the two scoring methods is in fact of self correlation.

3.9.4 Validity and Reliability

Studies of the cloze, for example, Ruddell (1964); Bormouth (1965) and (1967); Callant (1965); Crawford (1970), as a measure of native- speaker reading ability showed that the validity coefficients ranged from .61 to .95 between cloze tests and reading comprehension tests.

Studies on cloze as a measure of English as a foreign language proficiency have shown a validity coefficient of .71 to .89 between cloze and standardized English as a foreign language test batteries (Conrad 1970; Darnell 1970; Oller 1972b; 1972c; Irvine et al. 1974; Stubbs and Tucker (1974)). Brown (1983) has commented that such studies have relied heavily on validity coefficient which define the validity of cloze in general terms and do little to explain how and exactly what cloze is testing.

Studies made by Taylor (1953), Potter (1968), showed that the exact word replacement is practical and efficient, Davies (1978). Oller (1972) found that exact word and contextually appropriate words were better than exact word only for validating correlations. Oller (1972) found that the item discrimination for all texts was better with the exact word, i.e. the original word used by the author than with the acceptable word. On the other hand, the item reliability was better with contextually appropriate words. Oller thus recommends the use of the contextually appropriate words with foreign language learners. Haskell (1973) discovered that for measuring English as a foreign language proficiency, exact word methods were best. Irvine et al. (1974) seem to agree with Haskell.

Davies (1978) reported that correlation studies of cloze tend to

be contradictory. While people such as Oller (1972) points out, cloze tend to correlate highly with standardized proficiency tests, others such as Darnell (1968) and Mason (1972) reported low correlations with external criteria.

Cloze remains a controversial issue, there are striking disagreements on the issue of what cloze is testing. Oller (1973) suggested that what cloze measures is related to grammar of expectancy, as well as to short term and long term memory. His recent position (1979) is that the long range constraints are limited merely to cohesive devices. Farhady's (1980) arguments regarding the notion of what cloze is testing are formulated in his following statement:

> "A slight procedural change will ... change the whole purpose of the test"

Brown (1983) argued that Farhady's claims tend to ignore the large number of criterion-related validity studies which were conducted on cloze testing. Brown, therefore, took these assumptions further and formulated his research questions as follows:

(1) Does a cloze passage test cohesive devices and to what extent?

(2) Does changing the procedures for item selection change what the cloze passage is testing?

He addressed the first question by linguistic analysis and the second by statistical analysis and concluded that cohesive devices play a substantial role in what, the cloze passage, which he used, is testing. His findings support Oller's claims, that there are long-range constraints involved in what cloze is testing rather than

Alderson's (1979) finding that such constraints are of minimal importance (see Brown, 1983, for an elaborated justification).

Brown's experiment has also demonstrated that the role of the cohesive device does not change significantly when the procedure is altered. He, therefore, concluded that Farhady's (1980) claims are not substantiated, at least in terms of cohesive devices.

I found that research studies results on cloze testing in general is dominated by contradiction, therefore, to put it in Davies' own words, 'Cloze must be treated with caution'.

3.9.5 Conclusions on cloze

Cloze, as a testing device, has recently received a vogue. There exists a vast literature on cloze. For a recent extensive study and bibliography see Alderson (1978). Since its introduction as a testing instrument by Taylor (1953) views on cloze ranged from a test that is related more to low order skills (Alderson 1978) (lexical items and grammatical categories) to a pragmatic test, a test of overall proficiency, Oller (1979) related to all aspects of foreign language proficiency. On the basis of these differing conclusions, generalization is a bit of a problem. Cloze to me is hardly free reading comprehension, it involves an element of reading as either discrete grammatical or lexical based tests do (those realized in utterances or sentences). The nature of the deleted items, to me, are the only constraints that should decide the nature of the test, it therefore follows that any cloze test does not automatically qualify as a test of overall proficiency but rather a test of what is deleted. As Klein-Braley (1983) rightly point out that cloze should be subjected to very careful empirical scrutiny

since studies so far seem to be contradictory.

3.10 Dictation as a testing device

Dictation, like cloze, has recently received a great vogue. However, views on dictation ranged between a device that would measure little of language aspects to a pragmatic test, i.e. a test of overall proficiency in language. Below we will survey the literature to observe the status of this testing device. Traditional and current views will be brought together. My point of departure is Lado (1961). Lado viewed dictation as:

> '... it appears to measure very little of language. Since the order of words is given by the examiner as he reads material it does not test word order. ..., it does not test vocabulary. It hardly tests aural perception of the examiners pronunciation .... Spelling and a few matters of inflection and punctuation can be tested through dictation.'
>
> (Lado, 1961, p.30)

Rivers (1968), despite reservations on the technique, pointed out that dictation could be regarded as a useful device to test ability to discriminate among sounds, could be used as a test of the students' knowledge of combinations of letters and of structural elements. The reservations expressed by her were that testees do not pay attention to the meaning of what they write. She considered dictation a test of auditory memory more than a valid test of listening comprehension alone. But argued that it is a good technique as a teaching device.

Harriss (1969) argued that dictation as a testing device is both imprecise and uneconomical. By 1970 views on what dictation tests have changed radically. Oller (1971), Oller and Conrad (1971), Oller (1972), Irvine, Atai and Oller (1974), Oller and Streiff

(1975), Oller (1979) claimed that it is a test of overall proficiency. The studies cited above looked at dictation, cloze and some standardized tests such as the T.O.E.F.L. Irvine et al (1974) in particular concluded that T.O.E.F.L. provides little information other than that provided by dictation and cloze.

Valette (1977) seems to take a different view from that in Valette (1967). While she was of the view that language test experts question the effectiveness of dictation as a measure of advanced learners and considered taking dictation as a special skill, she (1977) is more able to consider dictation as a good measure of aural comprehension.

Alderson's (1978) research pointed out to contradictory findings that dictation was more related to lower order skills, vocabulary and grammar, than to the higher order skills. His research also confirmed previous findings, that of positive relation, high correlation between cloze and reading tests on the one hand and dictation on the other.

Attempts aimed at providing an interesting scoring procedure (Fountain 1974) that of scoring key words only. Another study by Moller (1982) was reported in the trial of his Communicative Proficiency Measure (CPM) using a similar scoring procedure. For practical reasons this subtest was not included in the final version of the CPM, because testees found it an easy task.

However, the fact that dictation is an auditory test is indisputable. The fact that it is integrative in so far as language skills involved in taking dictation are concerned may well be indisputable too.

The survey cited above points to three conclusions. Firstly, dictation is viewed as resources consuming and that different aspects of E.F.L. which could be tested through dictation may well be equally tested through more precise, objective and economical tests. Secondly, that dictation is more efficient as a teaching device than as a testing device. Thirdly, the fact that dictation tests used by Oller et al (1974) correlated positively with more than one aspect of testing English as a foreign language (vocabulary, grammar, cloze), and more the disagreement among language testers of the precise nature of the dictation as a test led Oller to consider dictation as a measure of overall proficiency.

However, more research on dictation is needed. Research questions may well be directed at the precise nature of the test as well as at providing scoring methods.

3.11 Communicative testing

The discussion of the notion of "communication" takes us back to the distinction drawn between 'language competence' and 'language performance'. Chomsky (1965) distinguished between the two in terms of an ideal native speaker-hearer viewpoint. He noted that the former is related to "the speaker-hearer's knowledge of his language" and the latter is to "the actual use of language in concrete situations". The term communicative competence, as found in the literature has been used to denote different things. To some, it means the ability to get a message across regardless of its linguistic accuracy. To others it relates to a set of abilities: linguistic, sociolinguistic and rules of discourse. Some also define the term as the social rules of language use. Duran (1984)

defined communicative competence as the control of all the functions which language may serve in everyday sociocultural contexts.

The notions, i.e. competence/performance, have also received frequent discussion by a number of language testers (J.B. Carroll (1961), Spolsky (1968), Jakobovits (1970)).

Moreover communicative testing was called for after new methods of language teaching methodologies, syllabus design viz. teaching language for communication have been inaugurated.

The need for language testing to catch up with language teaching have been expressed by writers in the field of language teaching. Wilkins (1976), for instance concludes that while people are experimenting with syllabuses, others should be attempting to develop the appropriate testing system that should accompany these new trends in language teaching.

Since then, serious attempts have and are being made to assess the feasibility of language tests based on communicative criteria.

Beginning with identifying what would constitute communicative competence. A number of proposals were outlined. Canale and Swain (1980), for instance, have been concerned with determining the components of communicative competence. The theoretical framework they proposed was mainly derived from the work of Hymes (1967, 1968), Halliday (1970), Widdowson (1978), Allen and Widdowson (1975), Wilkins (1976) and Morrow (1977). They isolated three factors which they considered as components of 'communicative competence'. These are:

Grammatical competence, sociolinguistic competence and strategic competence, for detailed considerations of the theory see Canale and

Swain (1980).

However, the major debate on communicative language testing concerns itself with the nature of the components of "proficiency" in language, in other words the construct and content of language tests. The nature of the theoretical framework on which language tests are designed has recently received more attention from language testers (Morrow (1979), Davies (1978), B.J. Carroll (1980), Moller (1982)). The notion of communicative language teaching has recently been reflected on the work of language testing. In Britain for instance, there has been a shift of emphasis in language proficiency testing from language "usage" to language "use". Significant work has now been completed in ESP testing (E.L.T.S., 1978 and Weir 1984).

However, despite the significant change in views of language proficiency, I find myself in agreement with Davies (1978) and Rea (1985) that communicative testing is still largely programmatic. There is no clear evidence as yet that these tests, i.e. communicative based tests, are better than the tests they are replacing in terms of either prediction of success, or ordering the students. What seems to be the difference is one of construct, one of adding another dimension to the definition of proficiency in language and that is communication, which implies the control of language in all sociocultural domains.

Over the last few years there has been concern as to whether the standardized test batteries such as E.L.B.A., E.P.T.B., actually measure proficiency (Moller, 1981). This concern is formulated in the following: student ability to engage in satisfactory

communication is not being assessed. In addition, there has been concern regarding validity with its four kinds (Moller, 1982).

To cast a look at the history of the communicative approach to language testing, our point of departure is Wilkins (1976). Towards the end of his book, Notional Syllabuses, Wilkins emphasised the need for testing to catch up with language teaching. In his conclusion he remarked 'we do not know how to establish the communicative proficiency of the learner'. He continues, 'While some people are experimenting with the notional syllabuses as such, others should be attempting to develop the new testing techniques that should ideally accompany it.

Two years later Munby (1978) came up with his book 'Communicative Syllabus Design'. This book attempts to solve the problem by designing a dynamic processing model (Carroll 1980) that starts with the learners' needs and ends with his target communicative competence.

The arguments put forward by the proponents of this view is an emphasis on language use. The argument they point out is that conventional language tests do not assess the use of language in a communicative situation.

Morrow (1981) specified one main characteristic feature of the communicative approach to language teaching and he says 'it enables us to make assumptions about the types of communication we will equip our learners to handle'. This assumption also applies equally to communicative testing. It, therefore, follows that what will be offered are tests of proficiency, at different levels, in terms of specified communicative criteria. Morrow concluded that 'there is

unlikely to be, in communicative terms, a single overall test of language proficiency but rather different tests in accordance with learners' operational needs'. In comparison between communicative tests and those which he referred to as conventional tests, integrative and discrete item tests, Morrow concluded that conventional tests do not give any evidence of the testee's ability to use the language. In addition there are certain features which are present in communicative tests but are not in conventional tests. Among these features is 'authenticity'. This claim has been taken further to be examined by Moller (1981) and Davies (1980). They concluded that communicative tests, in fact, lack "authenticity" as far as the resembling real communicative situations outside the testing room (for further details see Moller (1981) and Davies (1980)).

Moller (1981), Weir (1981) and Alderson (1981) agreed that Morrow's paper (1981) is an important contribution, but its shortcoming is it did not propose a straightforward definition of communicative testing, although it provided certain characteristics of the approach such as: communication is "interaction based, unpredictable, takes place in context, it has purpose and finally is behaviour-based".

Moller's paper (1981) involved proposals to define communicative testing on the basis of the following questions:

> How is language viewed and what is meant by language performance?
>
> What is to be tested?

His definition of language performance can be summarised as "an

assessement of the ability to use one or more levels of the system of the language so as to communicate, receive and understand the intended meanings of ideas and information by another speaker or writer and to be able to communicate such ideas and information to another speaker/reader".

The model proposed by Carroll (1980) of the communicative competence is based on Munby's specifications (1978). The specifications include:

(1)	Participants	Identification of participants
(2)	Communicative needs processor	purpose, dialect, setting, Target level, interaction, Event activity, instrument, ability, key
(3)	Skills/functions selector	Language skills
(4)	Specification	Test content

This model has been criticized for it neglects such pedagogical considerations whilst still identifying and specifying learners' needs (Carroll 1980). On the other hand, although the model has been carefully worked out, Carroll admits that it has got little empirical support in terms of scientifically collected and analysed data. One main advantage of the specifications is that they seem to offer a scheme for assessing performance both productive and receptive in a way which gives prominence to face validity. Face validity is not enough and as Criper (1981) rightly points out "What will be needed will be a programme of development and validation over several years which will deal with the real world of testing and needs rather than the hypothetical constructs of Carroll out of Munby" (Criper 1981 p.120). See (3.16) below for further criticism

of the E.L.T.S.

3.11.1 Characteristics of Communicative tests

(a) Emphasis on language use.

(b) Bias towards testing productive skills

(c) Use of rating scales for testing writing and speaking skills

(d) It can employ either discrete point items or integrative

(e) Emphasis on validity, particularly face and content, rather than reliability

(f) Proficiency is viewed as being effective communication

(g) It is concerned with proficiency in specific contexts and not in an overall sense

(h) It does not place a great emphasis on accuracy in the use of the levels of the system of the language, e.g. syntax or phonology

Examples of tests in current use are E.L.T.S. and R.S.A. in U.K. however they are still in the process of validation.

3.11.2 Descriptions of E.L.T.S.

The English Language Testing Service is an up to date system of language testing. The test was devised by the British Council together with Cambridge University local examination syndicate.

The test covers the language and language study skills needed for courses of study and uses the kind of texts and tasks one may find in his field of study. It measures the language skills of students whose native language is not English, who need to provide evidence of their English language proficiency. Its aim is to find out whether the learner's ability will meet the requirements of his/her course of speciality.

The test consists of two sections:

(A) A general section which tests listening and reading comprehension and is intended to test general ability in the use of English.

(B) A modular section tests language study skills used in reading, writing, listening and speaking and is related to a particular field of study. The modular section uses material from real books of the field of study.

There are modules of the subject areas of life sciences, medicine, physical sciences and social studies and technology.

The test has five subjects:

(A) General:

(1) Reading Comprehension

(2) Listening Comprehension

(B) Modular:

(1) Study skills

(2) Writing Tasks

(3) Interview

3.12 <u>Testing for specific purposes</u>

As an approach to the design of lanuage teaching and testing material, ESP has recently received considerable attention from syllabus designers as well as language testers. The general assumption underlying the construction of ESP material for instructional and (or) testing purposes is that language instruction should be directed to the construction of materials to meet the learners' particular needs related to their areas of study i.e. specialized instructions.

Strevens (1980) outlined the characteristics of ESP as largely

specialized instruction, developed to fulfil the utilitarian purpose of the use of a language in a particular occupation or academic setting (see 2. below).

By definition, then, specialized instructions may need specialized testing.

One major argument put forward on behalf of ESP testing and teaching material is that both have tremendous face validity. The argument put forward is that, for instance, a language test designed to assess students' proficiency, who are to embark on an electrical engineering course conducted in the medium of English, for instance, should look like a test of English for electrical engineers, i.e. input material for skills tested (communicative or linguistic) should be drawn from related subject matter. The second issue is that these types of tests may have diagnostic value, i.e. needs, or perhaps (requirements) analysis is centred on providing an answer to: what students will do with their language? proficiency to do what?

In answering the previous question, one is compelled to include in his test, only these activities his target population is going to deal with. This kind of testing procedure may have a good backwash effect, consequently may well be used for appropriate placements, i.e. areas of weaknesses may be detected and appropriate remedial programmes could be assigned.

In assessing needs, one allows for explicitness about what is expected of a student if he is to succeed in a particular occupation or academic setting to do with his English. However, the proponents of assessing needs for a particular discipline are met with two fold

problem.

a. Detailed needs analysis generate so many microlinguistic or communicative skills that it is impossible to include them in a test of an acceptable length (literally millions as B.J. Carroll acknowledged). It, therefore, follows that if traditional item analysis formula is to be applied on items for validation purposes, it should be applied with caution otherwise items which may be of importance may be eliminated subject to poor performance or extreme easiness.

b. The second major problem is with input material. Should they be authentic? i.e. taken unchanged from students' subject matter. If this is the case one should always differentiate between testing language proficiency to follow specific disciplines and testing the subject matter so that he does not end up testing for subject matter purposes.

3.12.1 Example of ESP tests

In the U.K. tentative attempts of ESP testing had been reported two decades ago. During the construction of the E.P.T.B., Davies made a gesture towards ESP in the specialised listening and reading tests. Gross differences were realized between 'science' and 'arts'. Specialists tests were tried but due to unsatisfactory results, they were dropped from the final version of the E.P.T.B.

More recently two serious attempts have been made towards the construction of ESP tests. The first one has been made by B.J. Carroll (1978) with the cooperation of the Examination Syndicate in Cambridge University. The second one has been made by Cyril Weir (1984) with the cooperation of the Associated Examination Board.

Both studies adapted two different models for the specifications of the tests. The first one was based on Munby (1978). The second one used an observation schedule adapted from Egglestone, Galton and Jones (1975) - the Science Teaching Observation Schedule.

Both attempts were aimed to do needs analysis, the first one is linguistic and communicative needs and the second one is needs analysis of events and interactions. The E.L.T.S. is the result of Munby's specification, the aim was to produce profiles arrived at through needs analysis. The attempt made by Weir aimed to gain an overall impression of what takes place in lectures, seminars and practical classes in different disciplines in order to come up with common activities and requirements among different academic disciplines.

The problem with information gained through frequency of occurrence is that the more frequent event or activity may not necessarily pose a problem. However the main advantage of these types of tests is one of diagnostic value. The E.L.T.S. and the T.E.A.P. type tests are based on specification of what the students have to do with their English. The value is that it could be judged, if the failure is to do with any particular area of difficulty, students may therefore be given appropriate remedial classes.

Another attempt of ESP testing was made in the U.S.A. by Erickson and Molloy (1983), but it is more of an experimental exercise than an operational test. The development of the test addressed one main issue: are there differences in performance on the test between native and non-native English speakers who are

doing engineering and those who are not?

The study reported a significant difference in performance on the test between native speakers and non-native speakers as well as a significant difference in performance on the test between engineering and non-engineering majors.

3.12 2 Conclusion on ESP testing

The need to test whether a non-native student is able to cope with and follow a particular discipline from a language point of view is necessary.

The mere dependence on speculative or theoretical framework for specification purposes could be dangerous. We have already seen an adaptation of a theoretical framework in the developments of a series of proficiency tests, that of Munby/Carroll type. It remains to be seen whether these series of tests are as valid as has already been assumed.

3.13 Measuring Spoken Proficiency

Speaking ability is the most difficult skill to assess with precision than any other language skill. Speaking is a complex skill requiring the simultaneous use of several different elements of language. Five elements are generally identified in the analyses of the speech process. These are pronunciation, grammar, vocabulary, fluency and comprehension.

The work on spoken production tests has been directed to provide a framework of scoring (i.e. method of scoring) either making use of analytical grids to make scoring reliable or testing the overall fluency by using an overall impressionistic scale.

In this review we will be concerned with some of the scoring

systems devised by writers in the field with some details on the FSI rating procedure. Madsen and Jones (1981) distinguish five categories of oral proficiency testing. All reflecting elicitation strategy and scoring technique. They fall along a continuum. At the one end, question types designed to generate communicative language; at the other end, techniques to facilitate evaluation of specific sub-linguistic skills.

3.13.1 a. <u>The 'Upshur' scoring method</u>

John Upshur has been concerned with oral communication tests of the type that makes use of the visual cues. The student sees a group of pictures, each identified by a letter. The student then is given a letter and must describe the corresponding picture. Upshur identified two scoring systems:

(1) Scoring the amount of communication:

Testees are given 36 sets of pictures and are required to describe one in each set. The total score represents the number of pictures described accurately so that listeners could identify the picture.

(2) Scoring the rate of communication:

The total score represents the number of pictures described accurately within the allowed time. This scoring system encourages active listening and helps the testees to realize that communication requires both speaking and listening skills.

3.13.2 b. <u>The Schulz Communicative Competence Scale</u> (1977)

This scale was developed by Rente Schulz to be used in the evaluation of 'free expression tests' using questions such as:

Describe this object?

The testee may say anything he wants about its size, use, etc. but he may not name it nor give a synonym for it. The response, then, is recorded and analyzed and scored later. As the name suggests the Schulz Scale provides a link with the question of the testing of communicative competence (Davies 1978).

Four factors are taken into consideration when scoring the 'free expression', these are: fluency, comprehensibility, amount of communication and quality of communication.

3.13.3 The work of the Foreign Service Institute (FSI)

The oral interview test of speaking proficiency devised by the FSI takes five factors into consideration but the proficiency rating is an overall judgement about the student's speaking competence. Scores are given as a single global score e.g. (S+3). The factors identified for the evaluation are the following:

Accent	foreign <——————————>	native
Comprehension	incomplete <——————————>	complete
Fluency	uneven <——————————>	even
Grammar	inaccurate <——————————>	accurate
Vocabulary	inadequate <——————————>	adequate

The standard practice among the FSI's experienced testers is to record the global score first at the end of the oral interview, and then to record a specific value on each of the five factors.

As can be observed all the rating scales mentioned above and so many others such as Clark four-scale system, Jakobovitz - Gordon and Bartz rating scales are attempts to provide a more reliable scoring system.

The presence of both the global score and the categories or

factor scores in the FSI scoring system led Clifford (1980) to investigate the question as to whether global scores are predictable on the basis of factor scores or not? He reported that scoring sheets for over 900 FSI tests were used as the data source for the study. These sheets consisted of both global FSI score and factor scores on the five skill categories. Clifford concluded that the first finding was that all the FSI factors do play a role in general language proficiency. All the factor scores contributed significantly to the analysis. In spite of this significant finding, Clifford reported that overall proficiency rating is not entirely predictable on the basis of factor scores. He, then, hypothesized that: 'Not all of the individual factors contributing to general language proficiency may have been accounted for'. By this Clifford is pointing to the fact that speaking skills will remain very difficult to assess with precision.

3.14 Language proficiency, unitary or divisible?

Perhaps the most disputed area in testing is that of the concept of 'General language proficiency'. The major question in the whole issue is 'whether or not all performance in a foreign language can be attributed to a single factor? Vollmer (1981) or as Alderson (1981) puts it:

> 'Is there one underlying proficiency in a second language, or are there several proficiencies, which are separately measurable and teachable?'

To these questions, three different views have been proposed.

It has been argued that language skills could be divided up into components similar to the elements of the system of the language, the position adopted by the discrete point testers, that of Lado and

J.B. Carroll. This position was referred to as the divisibility hypothesis. Within this theory it could be argued that knowledge of grammar, lexicon and phonology are different, and one is able to test them separately. In addition it is also possible to distinguish or even test or teach the four main skills of the language separately.

Theoretically speaking this approach is based on the assumption that there is no such thing as a single unitary language ability but a number of linguistic and non-linguistic competencies underlying language behaviour. The matrices given in Valette (1967, 1977), Harris (1969), Davies (1977) and Carroll (1980) identify these skills and components.

The second position is the one which sees the construct of language proficiency as an Indivisible or Unitary.

The proponents of this view, e.g. Oller (1979), argue that performances in a foreign language can be traced to one factor. Oller (1979) says:

> 'The construct of language proficiency may be more like a viscous substance than like a machine that can be broken down into component parts.'

He continues, 'this, referring to the hypothesis, cannot be excluded by pure logic'. The chief danger of this approach is that one may be led to believe that no matter what test one uses it will always rank students similarly. Decisions could therefore affect people's future.

The third hypothesis is that of taking any point between the first and the second positions. It was referred to as 'partial

divisibility hypothesis'.

Common sense suggests that people have different skills when it comes to the performance of the skills of the language. This can be equally applied to L1 and L2. 'Some people can speak and understand when spoken to, but cannot write an acceptable form of the language' Alderson (1981). Moreover, the matter has a pedagogical side. In the foreign language teaching situation, some skills such as reading are taught separately. Alderson (1981) argues that it is expected that there will be a variety of different proficiencies underlying performances on various tasks, having in mind the vast variety of sociolinguistic-settings in which it is possible to perform in a foreign language. However, a number of factor analysis studies were carried out to test these hypotheses. For further discussions see 3.18 below.

3.14.1 Factor analysis of language proficiency tests

The factor analysis technique as applied to language tests data is a statistical procedure normally used to quantify the relationship between the variables in the test (sub-tests), and to group them in relation to one or more factors. One of the major debates that has emerged from factor-analysis procedure has been the claim that there is a unitary factor of language proficiency which accounts for all variations in language tests. Oller (1979, 1983) presented data from different sources in support of his claim. However, the appropriateness of the statistical procedures he used were seriously questioned. Various factor analysis studies, for instance Palmer and Bachman (1981), Vollmer and Sang (1983), Farhady (1983), J.B. Carroll (1983), indicated the existence of more than

one factor underlying different abilities of proficiency in language and supported at least the partially divisible hypothesis, rejecting the strong claim made by Oller, that of 'the unitary competence hypothesis'.

3.15 The concept of validity with particular reference to validation studies of proficiency tests

Developments in language testing research have introduced and emphasised the concept of validity with its four kinds as important features of a good objective test. Almost every publication on language testing or even on testing in general has devoted some space to the illustration of the concept of validity. These publications, for instance, are particularly useful: Lado (1961), Cronbach (1964), Valette (1967, 1977), Harris (1969), Davies (1968, 1977), Ingram (1974), Heaton (1975) and Harrison (1983), although not in detail. With the exception of Davies (1968) the technique of item and test construction have received much more discussion and illustration, Moller (1982). It is worth noting that most of these books outlined above and in particular Lado (1961), Valette (1967, 1977), Heaton (1975) and Harris (1969) were mainly written as guides for foreign language teachers who are mainly concerned with the writing of achievement tests based on specific syllabuses. The validity of these tests will entirely depend on whether these tests are representative samples of the syllabuses or not in addition to what Moller (1982) said that 'the validity of these tests will depend on the validity of the syllabuses'.

Moreover the emphasis as well as the concern over the concept of validity was confined to the evaluation of standardised english

language proficiency tests based on the psychometric structuralist approach in testing. Other tests, based on other trends such as the communicative trend, for instance the E.L.T.S., is in the process of validation. Research is now being undertaken by the Institute for Applied Language Studies, University of Edinburgh, to investigate the concept of validity with regard to the E.L.T.S., an up to date testing system.

A very general definition of validity is the extent to which a test measures what it is supposed to test or measure, and how well does the test measure? There are four types of validation which can be applied to tests; all of these types attempt to answer the following questions.

What does a particular test measure?

and how well does it measure?

These types of validity are: face, content, construct and empirical, the latter has two subdivisions, predictive and concurrent validity. They can be established by statistical analysis. As for content validity, it can be established by linguistic analysis. Construct validity is established by the means of hypothesis formation and testing. It seeks to answer the question, how can scores on this test be explained psychologically? Cronbach (1961).

We are going to examine these types of validity in some detail and with reference to the validation studies of the standardized tests, i.e. E.P.T.B., E.L.B.A. and T.O.E.F.L. and others.

3.15.1 The concept of validity in relation to language proficiency tests

The concept

With the introduction of objective testing during the psychometric-structural phase of language testing, the notion of validity has received more attention from both test consumers and constructors. The studies of the concept of test validity are, in general, investigations into the extent to which a test measures what it is supposed to measure. Although the process of test construction necessarily entails the process of validation and despite the general concern of test validity, many tests are still published without any validity studies being reported (Moller 1982).

3.15.2 Kinds of validity

The concept of test validity and its types has received plentiful discussion in recent publications. These publications are particularly useful:

[Anastasi 1961, Cronbach 1960, Davies ed. 1965, Davies 1977 and 1978, Cronbach 1971, Palmer and Bachman 1981, Palmer and Groot 1981, Moller 1982].

Language testers have distinguished five types of validity. These are face, content, construct, concurrent and predictive.

3.15.3 Language proficiency tests and validity

Of the different types of validity, three types are most important and most related to the process of validation of language proficiency type tests. These are: content, validation, construct validation and criterion- related validation which is composed of concurrent and predictive validity.

3.15.4 Content validity

Content validity is largely linked with achievement rather than proficiency testing, the reason being that content validity for proficiency testing, where the universe to sample from is so vast, might be an impossibility to achieve. Teachers, past and present, are being concerned with the choice of test items relevant to what they have been teaching to their students. However achievement tests are sometimes classified as proficiency. In the field of teaching/ testing English for special purposes, for instance, content validity plays an important part in the construction of the tasks relevant to what candidates have been doing or expected of them to do. It is also at the heart of functional tests. The F.S.I. interview test and the P.L.A.B. test derive their content from tasks relevant to situations of the kind candidates will encounter, in these cases, occupational settings.

Investigating content validity in proficiency testing, is by definition examining the concept of proficiency as defined by the test constructor. Moreover examination of the domain of proficiency in relation to what is expected of candidates to do reveals the extent to which the test samples tasks and items relevant to future needs of particular occupational or educational settings.

Dissatisfaction with language proficiency tests which are linguistically based tests originated with the fact that these tests do not sample its content in relation to language use or more in relation to candidates' later use of the language. In addition, these tests are being criticized for the restricted definition of 'proficiency'.

Attempts have been made so that proficiency tests become more valid in terms of content and construct [The E.L.T.S.]. However, even with this test negligence of some aspects of 'proficiency' in language and the sampling for the test remain among the problems that would reduce the content validity of the test. In my opinion establishing full content validity of proficiency tests may well be an impossibility unless we find adequate answers to these points, the same questions may well be asked in relation to construct validity:

(a) What is language?

(b) What is proficiency in a language?

(c) What are the language skills involved in following a particular occupational or educational setting?

(d) What is the order of their importance?

(e) What should the input material for the test be?

Attempts have been made to establish content validity of proficiency tests through sub-tests inter-correlations, item validity and description of language and of proficiency. Some were theoretical justifications and others were empirical (see Davies 1965).

3.15.5 Construct validity

Until recently construct validity has received little attention from language proficiency testers. While Davies (1965) noted that construct validity may not be necessary for language proficiency testers to provide evidence for the existence of this type of

validity, he later (1984) pointed out that "what matters above all is construct and content validity". Others such as J.B. Carroll (1973), indicated that the only possibility of deciding on a test's validity is by examining its construct.

Again the consideration of this type of validity takes us back to the consideration of the factors mentioned in the previous paragraph, i.e. What is language? What is language proficiency? etc.

Construct validity of language proficiency tests has been and is being confined to the following techniques:

1. Difference in means and SD of native and non-native speakers' performance on a particular proficiency test.
2. Subtest-intercorrelation
3. Principal Component Analysis/Factor Analysis
4. More recently the multi-trait-multi method convergent discriminant validation study, has been used in investigating construct validity of proficiency tests (for details see Palmer and Bachman 1981)

E.P.T.B. (Davies 1965) reported scores obtained by native and non- native speakers and concluded that the E.P.T.B. does discriminate significantly between them. Differences in means were also reported for the T.O.E.F.L., the evidence obtained showed that the means were high for both test total and sub-tests. Sub-tests scores, it was reported, were maximum or near maximum for native

speakers. The study pointed to the fact that the tests were relatively easy for native speakers which is in line with what was anticipated.

3.15.6 The multitrait multimethod as a construct validation procedure

The procedure was first described by Campbell and Fiske (1959) and was first recommended for assessing construct validity of proficiency tests by Stevenson (1974). Palmer and Groot (1981) summarises its principles as follows. It is based on the assumption that a test score is a function both of the trait (the language skill tested or the level of language tested, i.e. [listening, reading, etc. or grammar, vocabulary, etc.] and of the method (the technique of testing used, i.e. grammar could be tested through multiple choice items or fill in the blank with the missing grammatical item being tested.

Under this method of construct validation, two types of validity are investigated. One is called the convergent and the other is known by discriminant validity of the test. The procedure assumes that tests of different traits may have low correlations between scores and this can be said to give evidence of discriminant validity. Evidence of convergent validity is obtained if one trait, for instance, grammar, is tested through different methods, one should observe high correlation between methods testing one trait or skill.

3.15.7 Predictive validity

Previous studies in predictive validity (Davies 1965, T.O.E.F.L. 1970, Maxwell 1965, Ingram 1973, Heaton and Pugh 1974), have all

used non-linguistic criteria. Perhaps by definition the criterion has to be something other than a purely linguistic one. The criteria used were normally end of term, session or year grades or exam results in the subject matter, i.e. academic specialities.

In the process of T.O.E.F.L. predictive validation studies, the Grade Point Average (G.P.A.) was used as the criterion against which T.O.E.F.L. was validated. The undergraduate and postgraduate education system in the U.S.A. uses the G.P.A. as the assessment scale. Five categories may be awarded for different kinds of performance. These are:

	Points	Marks obtained	Assessment
A	4	90 - 100	Excellent
B	3	80 - 89	Very good
C	2	70 - 79	Good
D	1	60 - 69	Pass
E	0	Below 60	Fail

This kind of assessment is widely used in the undergraduate and postgraduate education in Saudi Arabia. Maxwell (1965) in a large scale predictive validation study of the T.O.E.F.L. used the G.P.A. as a criterion. His study involved 238 students. The correlation coefficients for the entire group ranged from +.02 to +.58.

The second largest predictive validity was carried out by Davies (1965) as part of the process of the construction of the English Proficiency Test Battery (E.P.T.B.). Examination results were used as the criteria. The coefficients obtained varied from -.2 to +.7 with a total coefficient of .45. The total N was 208.

At the University of Leeds, Heaton and Pugh (1974) reported

findings which they regarded as confirming the view that language proficiency is important for academic success. The correlations obtained ranged from +.59 to +.66.

At the University of Edinburgh, Ingram (1973) and Howatt and Davies (1979) reported two large-scale predictive studies for E.L.B.A. The two studies have provided tables comparing E.L.B.A. scores with subsequent academic outcome end-of-term and end-of-session examinations. No correlation coefficients were reported by either studies. Ingram concluded that the higher the score on E.L.B.A. the greater the chance of success in examination results. Her results indicated that for those students scoring 70-79% on E.L.B.A. the success rates in the academic outcome were 68% and 90%. For students with scores between 50% and 69% on E.L.B.A. the percentage of success was 50% and 66% respectively. Those scoring below 50% were seriously at risk.

Howatt and Davies' study indicated that there is "a linear relationship between E.L.B.A. scores and academic success" (Davies 1984). The study pointed out that poor English is a factor in academic failure.

3.15.8 Conclusion on validity

Of the different types of validity, in my view, two types are of primary importance, one type is important and another type is the least important. The two most important types are content and construct validity. Content validity is the extent to which the selection of behaviour (language or otherwise) is representative to the larger domain. The universe of content defining language and proficiency in language can not yet be sufficiently described, and

as a result the demands of content validation cannot be fulfilled. Construct validity of a test is the extent to which the test may reflect a language learning theory or a theoretical conception of language. Predictive validity is the extent to which a test predicts future behaviour. Face validity is the least important type of validity since it only concerns the appearance of the test.

The choice of validation procedure depends on the use to be made of the test scores (Anastasi 1961). While content and construct validity enters into the process of constructing tests (the actual writing of test items), subsequent techniques of validation for establishing criterion-related validity may well be related to content and construct validity. However if a particular test is designed so as to predict future performance, it needs to be validated against the criterion of subsequent performance.

3.16 The notion of proficiency

The concept of proficiency in a second language has not yet been fully defined. The notion of competence or 'proficiency' has been explained in different ways, but all have in common a one fold query:

What does it mean to be proficient in language? Or as Spolsky (1968) puts it, what does it mean to know a language?

Jakobovits (1970) gave the following summary to the problem:

> 'The question of what it is to know a language is not yet well understood and consequently the language proficiency tests now available and universally used are inadequate because they attempt to measure something that has not been well defined.'
>
> Jakobovits (1970: p.75)

Due to our limitations in understanding the full nature of

language what Jakobovits, in effect, is saying is that content and construct validity of language proficiency tests are far from being achieved.

However in asking the question: What is it that we want to equip our second language learners with, that native speakers have but our learners do not? According to Stern (1983) proficiency in language could be viewed as a 'goal' and consequently may be defined in terms of objectives or standards.

It is widely accepted that complete proficiency in language (native-like proficiency), whatever its constraints and characteristics, can not be fully attained. However as Stern (1983) points out it may be desirable to keep that ideal goal as a point of reference.

During the past three decades significant work has been completed to offer language proficiency test constructors different schemes to specify the characteristics of proficiency in language. However, these specifications can be said to form a continuum ranging from purely theoretical base definitions to empirical base definitions. They are all attempts to specify the characteristics of language as a linguistic phenomenon and as a tool of communication in different sociolinguistic settings.

Some define proficiency as linguistic competence thus, one defines proficiency in language in terms of the skill-aspects approach (J.B. Carroll 1961). In this model it is assumed that four types of language proficiency are employed in each of the four macro skills. These aspects/levels or components, as different people refer to them, are knowledge of phonological and orthographic rules,

knowledge of grammar (syntax) and knowledge of lexical items.

Others give a more behavioural definition to proficiency. On this dimension proficiency is described as language activities which are described in terms of detailed inventories of language items, situations, topics, speech acts and functions. The Council of Europe defined proficiency in more concrete details of syllabus items designed for specific groups (van Ek 1975).

Another definition of proficiency is taken from Munby (1978) and formulated in the construction of the E.L.T.S. test that of the enabling skills. Now, the E.L.T.S. is supposed to be a proficiency test designed to answer questions such as can testees follow lectures on different subject matter? Will they be able to read and comprehend texts written in English? Not only that will they be able to understand relations within sentences or understand grammatical cohesion? We may well acknowledge that such micro skills underlie the ability to follow lectures and comprehend materials, among other things of course, but should they be the main dependent variable in a test aimed at measuring proficiency?

We do not know that these enabling skills are equally important or some are more important than others. We might also not understand their nature. Another point made by Hughes (1986) has to do with the extraction of texts for testing purposes. It is not clear whether we will find texts which will include all these enabling skills or even a random sample of them.

Canale and Swain (1980) interpreted proficiency as 'communicative competence' which involves, as was proposed in the theoretical framework they advocated, four areas of knowledge and

skill. These are sketched briefly below as follows:

a. Grammatical competence which includes mastery of the language code thus concerned with language features such as lexical items, pronunication and rules of sentence formation. It seems that this area is based on J.B. Carroll's definition of proficiency (1961).

b. Sociolinguistic competence which involves ability to use language appropriately in different sociolinguistic settings and contexts, with emphasis on both language form and meaning.

c. Discourse competence which includes the ability of how to combine and interpret forms and meanings to achieve a unified spoken or written text through the use of cohesive devices to relate sentences and utterances to forms and coherence rules to organize meaning.

d. Strategic competence which includes verbal and non-verbal strategies to convey the intended message and to compensate for breakdowns in communication.

As can be seen from the categorization above this theoretical framework is very broad. It is not obvious which area(s) of its elements is less or more important than other(s). Another point which is more related to the language acquisition aspect than to testing of how these areas or components would interact at different stages of language acquisition. It is also not clear how one would score a test in terms of the last dimension i.e. strategic competence.

Oller (1975, 1979) has proposed a theory that proficiency in language is unitary which he called grammar-based expectancy. This

theory has put forward a two part justification. The first is theoretical, related to the definition of language, that language forms a whole and is not unrelated segments and components when it is used for communicative purposes. The second is empirical and is based on the interpretation of test data viz. correlation and factor analysis. This second argument may not be plausible for several reasons. First that the number of factors one would get depends on the method of analysis. Principal component analysis simplifies data. It reduces it to one solution and therefore it is likely that only one factor emerges from the intercorrelations. Varimax rotated solution on the other hand looks for as many factors as possible. Another reason, the one of correlation between tests could be either attributed to the characteristics of the sample or the tests themselves. Language tests, it was suggested by Cummins (1979, 1980), have certain academic or cognitive aspects.

However the type of tests advocated by Oller i.e. dictation and cloze lack another quality of language use corresponding to what has been termed as communication. Oller's definition of proficiency will be discussed in some detail below.

My own view of the interpretation of the notion of proficiency is set out below.

When interpreting the concept of proficiency we should not think of it as an abstract term but we should consider and relate the consequences of the interpretation to language teaching which is an integral part of testing. Given the complexities of language it would be more appropriate to consider proficiency in language so as to consist of multiple categories containing theoretical conceptions

of language and practical conceptions related to the actual use of language. The theoretical position I advocate is that proficiency in language is different linguistic abilities more related to accuracy and fluency and better described in relation to the question proficiency in language to do what?

It is therefore my view that if we want to have a kind of meaningful feedback, proficiency tests should consist of parts assessing linguistic competence by one of the many ways which had been researched over the last three decades and other parts that would give some kind of indication of the ability to use language. In conclusion we should remind ourselves that tests are only representations of proficiency, given that test designers work under constraints such as time and resources, proficiency measures must approximate since it is unlikely that we can administer a test which will assess competence or performance in language covering all the specific situations in which testees would operate directly or indirectly.

3.16.1 Traditional aspects of assessing proficiency

In a fairly recent sketch of the developments of language proficiency testing, Spolsky (1978) distinguished two types of language proficiency tests belonging to two different stages. The first is the discrete point approach and the second is the integrative approach. However it looks evident that these different trends in testing have evolved from the new developments in language teaching. The emergence of the third stage psycholinguistic-sociolinguistic makes an essential assumption that language skills are best evaluated in context, in addition emphasis on language as a

communicative tool.

Below we are going to look at these types of assessing proficiency in some detail.

Following this line of argument i.e. that testing approaches emerged as a result of the influence of teaching we could link the discrete type of proficiency testing as being related to the discrete type teaching methods such as the behaviouristic or the structural approach to language teaching. Integrative tests may have emerged as a result of the influence of the cognitively based teaching methodologies, for example, notional/functional approach.

3.16.2 Discrete proficiency testing

Although the term 'discrete' has been well taken in the literature of language testing there is no clear definition of what kind of items would constitute a discrete item. The general assumption is that discrete item tests assess one particular feature or segment of language, be it grammatical, lexical or phonological. However, there are degrees of discreteness and items could be classified from highly discrete to integrative. An uncontextualized item, therefore, which requires the testee to choose a synonym of a lexical item from a list of distractors could be regarded as highly discrete.

A further assumption of discrete point testing is that language levels tested are referenced by a standard variety of language whose rules determine the level of correctness. Consequently proficiency as a measure is referenced to that variety.

The design of all types in general involves certain steps as construction, trialling and revision. Depending on how effective

the initial analysis had been, tests may result in reasonably reliable tests.

However the validity of the discrete item tests has been called into question. Discrete item tests break the components of language apart i.e. the complexities of language are analysed into bits, levels, skills etc. and are tested independently of each other. It is this theoretical basis that has been seriously questioned. The arguments put forward against the discrete approach to the construction of tests are that language is not a set of unrelated segments and components, it forms a whole so the segments should be integrated and tested in combination of one another. In addition that language as a tool of communication is contextualized and the purpose of language learning is usually communicative so what should be tested is communicative ability and not the form(s) of language (Davies 1978). Such criticism does not imply that there is no place for discrete item tests in proficiency testing or that integrative types are better. On the contrary, integrative testing, as we will see later has its problems too. Discrete item tests in my view are still one of the better alternatives in the developments of tests for diagnostic purposes for instance.

However, in spite of the objectivity of a discrete item test as well as its practicality, its shortcoming stems from the fact that answering individual uncontextualized items i.e. regardless of their function in communication, may not be of great value. It is not an indicative to testees ability to use language since the testees' role is limited to the recognition of formal knowledge of language at the surface structure level. One additional point is that

crucial properties of language are lost when its parts are isolated (Oller 1979).

3.16.3 The notion of integrative proficiency testing

The integrative view of language proficiency testing may be best described by comparing it with the notion of discrete item testing. The integrative view of language proficiency contrasts with the discrete point approach in the sense that proficiency is best assessed by looking at language use requiring integration of a number of sub-skills and components of language use. Oller (1979) has described a 'pragmatic proficiency test', a term he used to describe a kind of integrative proficiency test as follows:

> 'any procedure or a task that causes the learner to process sequences of elements in a language that conform to the normal contextual constraints of that language and which requires the learner to relate sequences of linguistic elements via pragmatic mapping, to extra linguistic context.'
>
> Oller (1979: p. 38)

The citation above, according to Oller, implies that pragmatic proficiency tests involve two constraints. First, that processing language by testees on these pragmatic tests must be constrained temporally and sequentially. This may imply that testees would process sentences in tests such as cloze and dictation, for instance, as meaningful units with some kind of an overall communicative meaning. The second constraint is that such tests use contextualized language resembling the natural occurrence, of language ouside the testing context. However it is not quite clear how a pragmatic test of the cloze variety would resemble the naturalness criterion of language use since it is realized in an

artificial context far from resembling actual language use.

3.16.4 Problems with integrative tests

Cloze and dictation are two types of the best known examples of integrative tests. Other types include written composition, reading comprehension requiring inferences and oral interview. These tests require testees to process language through coordination of a number of sub-language skills. They attempt to use many segments of language simultaneously, different levels of the language and perhaps more than one skill.

If cloze and dictation are contextualized in the sense that different aspects of language are tested in combination with one another, because utterances or sentences are interrelated, they are, in fact, decontextualized from the normal settings within which language is actually used. These tests, therefore, are inadequate in the sense that they do not tap and assess the testees ability to actually use language in communicative contexts.

Oller (1975, 1979) reported that dictation is as valid and reliable a test as any other listening comprehension test; that it represents a random sample of contextualized language. This conclusion is based on Oller's empirical research. However, there are disadvantages with the procedure. First a decision has to be made as to what would constitute an item. Second a procedure or a method of scoring has to be decided. Scores on dictation depend on several factors:

a. ability to understand the material being dictated

b. ability to write down what is understood

c. and, sometimes, a mechanical ability, that of speed

(Farhady 1983)

We might not be able to judge with precision whether a particular score is as a result of problems to do with comprehension or ability to write with speed. The old question of validity therefore persists. Does dictation appear to measure what it purports to measure?

Because cloze is a more controlled (Farhady 1983) task i.e. a cloze test has uniform items throughout the passage in contrast to dictation. The problem with cloze is not as severe as those with dictation. However it is a stated fact that we do not have a clear idea of what cloze is actually measuring.

3.17 Current aspects of assessing language proficiency

It seems logical to assume that the fundamental reason for the 'renewed' awareness of assessing language as communication stems from the fact that existing methodologies of assessing language proficiency do not adequately deal with the communicative use of language in appropriate contexts. The renewed awareness has focused attention on assessing the testees ability to handle the communicative functions of language rather than the ability to manipulate the linguistic forms of language only.

This renewed awareness has two fundamental features. The first is that language should not only be viewed as a linguistic phenomenon which consists only of forms but as a tool of communication. The second is that direct tests are most revealing as measures of testees language proficiency, i.e. their ability to use language in communicative situations.

There are advantages and disadvantages in this approach, i.e.

assessing the communicative competence of testees. We will begin with the advantages. The first is related to language teaching. Language teaching should aim at developing the communicative competence of the learner rather than linguistic competence. The second is that the objectives of learning as well as student needs should be central to the planning of the specifications of tests.

What this means is that learner-centredness (Alderson 1986) is viewed as crucial in the design of tests. This allows for explicitness with regard to students needs. Needs analysis forms the basis of the specifications of the content of tests aimed at measuring proficiency.

The major argument for a communicative-oriented approach to language proficiency is linked with face, content and construct validity. It is commonly claimed that communicative oriented tests have face validity and higher content and construct validity than linguistically based tests. However in view of the disagreement of the theoretical bases of communicative competence (Hymes 1972; Morrow 1977; B.J. Carroll 1980; Canale and Swain 1980) it is not quite clear what are the characteristics and therefore the right specifications for test format and content of communicative competence. Generally speaking it is accepted that natural communication is dynamic, interactive, unpredictable, purposive, authentic and behaviour based (Morrow 1977). The question to be raised is how these features of naturalness of communication identified above can be met and incorporated into the specifications of tests given that one usually works under other external constraints such as time and resources.

It should be noted that the backwash effect of any testing system should be of paramount importance.

While I am in agreement with the view that communicative based tests have considerable face validity, there are so many problems associated with these tests. First, tests are measures and measures are normal scored. Scoring of tests of the 'direct' nature is highly subjective and requires training. Second, in educational settings the assessement of large numbers of students might prove problematic. Third, it is unlikely that an assessment of communicative competence, given the limitation of time and resources, necessarily reflects all the sociolinguistic diversity of contexts and situations since the number and nature of language functions are so complex and uncountable. The most difficult problem that theories of communicative competence encounter is the definition of proficiency levels, i.e. how a particular performance is to be assessed.

Communicative-oriented tests are mostly concerned with appropriacy rather than accuracy, therefore, what are the set of definitions that correspond to particular language performance. Given that communicative proficiency and even general proficiency in language forms a continuum ranging from no ability at all, no communication, to native speaker ability, it would prove difficult to incorporate precise and feasible description of performance to realize all language functions, given the enormous variety of sociolinguistic settings in which it is possible to distinguish different kinds of performance. Another problem is to do with the norming groups, although one would argue that communicative tests

are criterion- referenced tests, the question to be raised is if one wants to establish proficiency levels how that is to be done. Can it be done with reference to native-speakers of the language or non-native speakers? If we choose non-native speakers then which non-native speakers should one use?

The problems, be they theoretical or practical in assessing communicative competence are challenging. Constructing tests solely on the basis of the communicative criteria of the naturally-occurring communicative contexts, for instance, may be regarded as inappropriate on the basis that this approach ignores other aspects of language proficiency. After all what are language tests? What are they for?

Language tests as described by Davies (1986) are measures and samples. Measures in the sense that a score on a test should be interpretable and samples in the sense that they represent language as behaviour whether linguistic or communicative. It could be argued that all language tests, whether of the discrete, integrative or communicative nature are samples of language. Whether they are representative samples of the situation or which they are intended is another issue. It is actually an issue of content validity of the respective test in relation to its use. There has been a demand for tests to be more explicit about different needs of different testees. This can be said to provide an advantage since the concern has been that purely linguistic based tests are not the type of tests that students need or they are too general to be of value for different groups of students. The implication of this is that tests should be differentiated in content according to the purposes of the

tests. Without commenting on the construct of the E.L.T.S. test, it can be said that it differentiates in content among different disciplines. So far as tests are measures, the question may be asked whether proficiency tests whatever their characteristics, make an empirical assumption regarding what non- native students will have to be able to do with the target language. So tests are measures in this way relating present proficiency in language to its eventual uses.

3.18 Unitary competence hypothesis

Is it plausible?

Oller (1979, 1983) argues that high intercorrelations among measures of proficiency in language and factor analysis of the principal component variety show that language proficiency can be regarded as a unitary skill. The argument is based on numerous empirical data which show substantial correlations between different measures of language proficiency, of the discrete variety such as tests of grammar, vocabulary and of the integrative variety such as cloze and dictation. He advocates that the high correlation is an indication of the existence of an internalized grammar. He also uses data belonging to the simple principal-component analysis revealing the existence of a general language proficiency factor. It seems that the issue has two facets, one theoretical, the other statistical. We will look at these points in turn.

Language as human behaviour is so complex that it interacts with so many other mental abilities. Looking at language from the point of view of skills, i.e. macro language, skills reading, writing, speaking and listening, we would expect that people have different

skills; some people can speak and understand language but cannot read or write, this is equally true of L1 and L2. Other people can read language but cannot speak it. If there is a general language factor why then, in our tests, do we test more than one skill? I think the assumption is, given the varieties and complexities of language, that it is expected that students will perform differently, in which case it can be said that underlying each skill is a different type of proficiency. Common sense therefore suggests that we do not need a statistical package to prove the obvious (Alderson 1981).

Relating this to the teaching of the different language skills it is assumed that these skills are distinct areas of performance. This is particularly noticeable when we direct our teaching or focus it on one skill, speaking for instance, we implicitly assume that speaking is an independent skill of language and can be taught and measured separately, and at one point of time we may say that our learners are able to speak the language more effectively than they can write it for instance.

We now turn to the empirical research viz. principal component analysis and correlations. We will begin with the principal component analysis. Principal component analysis as applied to test data looks only at variance structure rather than the covariance. It is therefore likely that the procedure simplifies the data in order to reduce it to one solution only. Therefore it is expected that only one factor emerges from the data. Maximum factor analysis, based on one of the methods of rotation, varimax for instance, looks at the covariance structures within the data; it

therefore looks for many factors underlying the data. In our present study, see chapter six below, our data was analysed with the two different methods, principal component analysis and varimax rotation. When the data was factor analysed using principal component analysis, one general factor and another smaller factor emerged from the data. When the data was factor analysed using varimax rotation two significant factors emerged. It is therefore required that one should use maximum likelihood factor analysis. Maximum likelihood factor analysis allows for several factors to emerge, see Farhady (1983), J.B. Carroll (1983), Vollmer and Sang (1983).

It is my view that if we disregard correlations and the factor analysis, and look at the problem from a practical level there is no one best test, we must give our testees a variety of tests. By this we will ensure that our decisions may be regarded as more appropriate given the complexities of language. We conclude by asking why are we interested in E.S.P. testing? Is it not the case that it is an implicit assumption that there is no general factor but there are several.

If language proficiency were completely unitary, which in my view it is not, it would be possible to use only one kind of test which is valid and reliable and forget about all the time and resources which have been invested in the development of an understanding of the notion of 'proficiency'. Until now there has not been a comprehensive understanding of the constraints and characteristics of proficiency.

3.19 Implications for the design of proficiency tests

The foregoing arguments concerning the notion of proficieny have pointed to different views of the concept which can be described in such a way as to form a continuum. At the one end and at the other, two extreme views were presented. Neither of them in my view is acceptable. The first sees language proficiency as a linguistic phenomenon, therefore testing the linguistic forms at the surface structure level is an indication of testees competence in language. The second is the specifications of test items in terms of those features of the communicative use of language in exact terms, viz. interaction based, "unpredictable, purposive, authentic, behaviour based" (Morrow 1977) or the Munby-like model (1978).

These extreme views, in my view, are neither desirable nor acceptable.

There are a number of issues that a proficiency test constructor should be concerned with. These are:

a. The influence of the test on the planning of teaching i.e. working for positive backwash effect.

b. Content validity with reference to the needs of the learners, and the language domain, i.e. adequate representation.

c. Method(s) of testing and scoring in relation to resources.

The main principle of a testing programme where teaching is offered is that our test construction should have a positive influence on the teaching situation. We have already seen an extreme example in this particular case study, how the E.C.L. prevents instructors from doing anything useful with cadets.

Backwash effect refers to the effect of testing programmes on teaching practices. Instructors in the Saudi Air Academy are teaching towards the E.C.L. i.e. because they are familiar with the content and format of the test, they are teaching their cadets according to what is in the test. In constructing a test one should ensure that his test contributes effectively to teaching practices, and does not distort them, i.e. the pedagogical function of tests should not be ignored.

The point of content is related to what one might call proficiency in action. Tests aimed at assessing proficiency to follow a certain academic or occupational setting must incorporate within them these skills relevant to eventual use. Content should also be motivating, materials are stimulating, instructions are explicit. It might even be desirable to give some suggestions with regard to answering tasks. Sufficient time should also be given, not only to complete tasks but to revise them as well within the available resources of course.

Methods of testing

One of the functions of proficiency testing is that test scores have major effects on where to place students in a language programme, for instance, or whether to admit them to a particular English medium course. If a test therefore is compiled solely of either discrete point or integrative tests some of the students and because of their familiarity with the procedures will perform better or worse than their 'true' score and may be placed in inappropriate classes. It is therefore unlikely that one type of test will reflect any optimum assessment of those complicated language

processes and it is likely that varieties of tests may give a better judgement.

3.20 The State of the Art (Summary)

In teaching as in testing there has been a shift of emphasis from one approach to another. The tendency of such changes is one of questioning the validity of previous method or methods.

In the field of language testing Spolsky (1975, 1978) identified three approaches. These were the prescientific, the psychometric structuralists and the sociolinguistic psycholinguistic test.

Influenced by the old grammar-translation method, the prescientific method paid no attention to the basic virtues of a good and sound test, i.e. reliability and validity. The form of testing was one of translation or composition writing or both. Such tests were scored subjectively.

In the second stage teaching and testing methodologies were substantially altered. As a result of the inauguration of the audio-lingual theory and structural linguistics, the development of objective tests dominated all psychological tests and language tests are no exception. The theory of testing was later known by the name of psychometric-structuralist. The basic assumption of the theory stems from defining language. Language according to Lado is a system of habits which involves form and meaning at the different levels of surface structure beginning with the smallest unit phoneme and ending with the largest unit of structure sentence. The type of testing associated with this trend was known as the discrete-point approach. Each test item was constructed on one point of language: grammar, vocabulary, phonology.

However, this approach to language testing has later been criticized. The fundamental criticism to the approach is that testing individual and isolated items regardless of their function in communication may not indicate the testee's ability to use the language. As a result J.B. Carroll (1961) suggested using types of tests which focus on the total communicative effect of the message rather than the sentence discrete components. He called these tests 'integrative'.

Farhady (1983) wrote:

> 'This was the start of the third and most recent shift in language testing'.
>
> (p.312)

The proponents of this view, among them Oller, noted that integrative tests in contrast with discrete item tests assess the skills which are involved in normal communication and they are valid instruments in so far as assessing the overall language proficiency of second language learners.

Examples of these types of tests are cloze, dictation, listening and reading comprehension tests and oral interview. It should be noted that surveyors of the historical developments in language testing when exemplifying each period do not mention any standardized tests of the last trend in language testing particularly cloze and dictation.

In comparison between discrete and integrative tests which form the second and the third phases of language testing development Farhady (1983) wrote:

> 'The shift from discrete-point to integrative tests is not yet fully accomplished, so they still tend to compete as separate avenues of testing'.
>
> (p.317)

It should also be noted that most of the research studies conducted on the development of integrative tests have used well-known discrete-point tests as the criteria to validate the former. Examples are Alderson (1978) used the E.L.B.A. and Moller (1982) used the E.P.T.B.

Farhady (1983) quoted some findings of several studies including his own to prove that basically there is no difference between discrete-point and integrative tests in what they measure. He commented on the findings as follows:

> 'However they (correlation studies) also indicate that the general language proficiency factor "whatever it might be" present in the cloze, dictation and reading comprehension sub-tests, "integrative" is equally evident in the grammar and vocabulary subtests (discrete-point)'.
>
> (p.318)

Correlation studies between discrete-item tests (vocabulary and grammar) on the one hand and integrative tests (cloze, dictation, reading comprehension and listening comprehensioin) on the other indicate the following correlation coefficients:

Rand (1976	.63, .70, ,62, .63
Oller and Inal (1975)	.63, .58, .61, .63
Oller (1972a)	.83, .79, .81, .80, .73, .79, .80, .69, .80 .73, .83, .77, .78, .71, .85, .76, .73
Bowen (1976)	.71, .70, .69, .50
Farhady (1978)	.87

3.20.1 In Conclusion

However, because we are going to use the scores of the test under development to make some meaningful decisions with regard to students' performance, i.e. (placements and proficiencies purposes) a pragmatic approach has been adopted. The situation might in fact dictate the approach one should work within in the development of his battery, not only this but language as a human behaviour is a complex ability and despite the high correlation findings quoted above it would be far from adequate to work within one test theory. We conclude by quoting Davies (1978):

> 'tests will continue to use a combination of discrete-point and integrative items ...'
> (225:1978)

The implications of this are if a test is constructed solely of discrete items or integrative sub-tests some students may appear disadvantageous on the basis that, assuming they need remedial English they may be placed in inappropriate classes. Adding to that it is very unlikely that one single type of test accounts for all segments underlying the complexities of language and proficiency in language.

CHAPTER FOUR

Test Design: Type of test needed, content and methods.

4. Argument

4.1 Type of test needed

A survey of the language testing literature reveals that several types of tests have been designed to meet different requirements and to supply the necessary information with regard to the purpose to which they have been designed, i.e. according to use. Achievement (attainment) tests, proficiency tests, aptitude tests and diagnostic tests are all possible types of tests in language and are distinguished in terms of three criteria, these are: content, time and use, i.e. purpose. (See Davies 1977 for detailed discussion.)

All these kinds of language tests, together with some others such as progress tests are discussed and exemplified in the major publications on language testing. With regard to testing English as a foreign language, these publications are particularly useful: (Heaton 1975; Valette 1977; Davies 1977; Cohen 1980; Harrison 1983)

By considering first the characteristics of each type of language tests cited above, we could arrive at some conclusion with regard to the type of test needed by the R.S.A.F. to be used for the flying cadets at the "K.F.A.A.". Starting with the function of achievement tests, they are concerned with assessing what has been learned of a clearly defined syllabus. Testees are assumed to have been trained within a unified curriculum. Thus, in the non-language field in Britain, for instance, the "A-level" and the "O-level" are typical achievement tests. The prime consideration of achievement tests is to indicate how much of a defined syllabus has been learned. They

can also be used as an evaluative instrument to provide some feedback not only in terms of students' performance on the tests but also to provide useful information about the effectiveness of teaching methodologies employed by teachers. While achievement tests look back on what has been learned of a defined corpus, of a language for instance, proficiency testing on the other hand concerns itself with evaluating present abilities in relation to future needs. Thus, in the language field, the T.O.E.F.L., the E.P.T.B. and the E.L.T.S. are types of proficiency tests designed to provide some prognosis and to evaluate proficiency in English as a second language. Although they differ in their interpretation of what constitutes proficiency (see chapter three) they were all designed so that they provide an answer for the following question:

> 'Does a student have enough English to enable him to follow a certain speciality given in the medium of English?'

Language proficiency tests, therefore, are concerned with measuring a student's control of the language in the light of what is expected of him (to perform) in terms of the tasks imposed on him by a particular job or a particular course of study.

A language proficiency test has been described by Davies (1965) as follows:

> 'A language proficiency test seeks to measure the adequacy of a subject's language control for something: this something is not, of course, usually a language course but any sort of academic study or, indeed, any sort of occupation which involves the use of the language.'
>
> (Davies 1965, p.57)

A proficiency test may or may not be based on a known syllabus but

it is designed in such a way that it may make predictions of future behaviour. It is used to evaluate present abilities in relation to future performances. However, if a proficiency test is based on a unified syllabus and testees were all trained on that same syllabus, the test can be labelled achievement. An achievement test, therefore, is a kind of proficiency testing.

In order to specify the type of test needed by the Academy let us briefly consider the Academic setting at the Academy. In addressing the following question:

> Why is there a need for an English language test? and what type?

The King Faisal Air Academy (K.F.A.A.) is like any other English medium institution. It trains cadets to be pilots in the R.S.A.F. and uses English as the medium of instruction in subjects such as: Aeroengines, Aerodynamics, Navigation, Meteorology and all the other related disciplines to Aviation Sciences. The student is required to show evidence of proficiency in the English language either upon entry or after a period of training in the English language, not later than 52 weeks from the entry date. English language training is provided but admission for subsequent training is not granted. A student has to demonstrate that he will be able to cope with subsequent training given in the medium of English.

Therefore, in order to meet the K.F.A.A. need for a prognostic test which would measure the suitability of a cadet's English for studying in the medium of English, i.e. Aerosciences and flying training, it was clear that the type of test to be designed is a proficiency test.

4.2 Principles of test design

A proficiency tester, planning to produce a proficiency test, must right from the beginning identify what 'constitutes' proficiency in a language in relation to its uses. He must also specify the criteria against which he will validate his test, and must also be concerned with practical matters, such as ease of administration and scoring.

4.3 Towards a theory of proficiency

Arguments

Currently there are two main approaches to the construction of language tests. On the one hand there is the "linguistic" approach, thus a test constructor starts off from the theory that language can be analysed into language levels, i.e. proficiency is defined as having a competence over the system of the language. Thus, language is analysed into several components, these are: phonology, grammar, lexis. Within this approach, the complexities of language are analysed into segments. A test constructor makes an assumption that knowledge of the elements of a language is equal to knowledge of the language. Therefore, the notions of construct and content theories require that language be broken into elements. Language, then, is tested in terms of its elements - phonology, grammar, lexis, and skills - listening, reading, writing and speaking. However, althogh this theory of language testing emphasises the importance of reliability and validity and an extensive use of objectively scored tests, its shortcoming, or the argument put against this theory of language testing is as follows:- The first is that language as it is used in communication forms a whole, language levels are related

when they are used in an actual communicative situation, so they have to be tested in combination with one another. The second is a language is learnt for a purpose, the purpose is usually communicative, so what needs to be tested is communicative ability rather than the ability to recognize the system of the language separate from one another (Davies 1978).

The argument presented above is summed up as follows: Testing the linguistic competence, i.e. formal knowledge of a language is not indicative of a testee's ability to use the language for communication purposes. Briefly, linguistically based tests were considered inadequate measures on two grounds. Firstly they do not measure the productive skills directly, and a high score on a receptive skill test may or may not show that a testee can express himself adequately in either the spoken or the written forms of the language. In this respect, Moller wrote:

> 'A good score on a listening test may or may not indicate that a subject can express himself adequately in spoken English on the assumption that each type of language test is in fact also measuring a general language proficiency factor, it would be reasonable to expect an adequate performance in spoken or written English, particularly from those with a percentile rank of 85 or above. Similarly we might expect that those below the 15th percentile would not be able to speak or write any English with any degree of adequacy. But we can have no meaningful expectations of those ranked in between.'
>
> (Moller, 1982, p.22)

Secondly, linguistically based tests do not require subjects to perform tasks relevant to what they will use the language for.

There is no doubt that these two approaches to the devising of language tests discussed in chapter three and outlined briefly in

the previous paragraphs "mesh" (J.B. Carroll 1983) with one another.

Communicative testing as an entry point of test design is only one dimension of proficiency in a language. It need not be the sole definition of language proficiency. On the other hand, linguistic competence is as important as ability to communicate and it may well be a prerequisite for communication forming the first element in language proficiency.

Having established the dimensions of proficiency, the second question which lends itself to investigation is:

> What would constitute proficiency in English for a cadet at K.F.A.A.?

From the arguments presented in the previous paragraphs we concluded that a cadet at the K.F.A.A. is required to provide evidence of his proficiency in the English language since he is following an English medium course. Not an English language programme, of course, but a speciality taught through English. That is aerosciences and flying training.

We might therefore conclude that what constitutes proficiency in English for our cadets is related to their success in aerosciences. Certainly English is only one contributing variable to successful achievement in his speciality but an important one no doubt. Thus, an English proficiency test must have a predictive power. It should tell us beforehand who and who is not able to follow an English medium course.

4.4 Validity Criteria

It was therefore decided to use two different criteria against

which to measure the validity of our test:

1. Instructors' estimates of their cadets' English proficiency.

2. Cadets' examination results in their subsequent academic subjects.

The first criterion has been criticised for being an unreliable measure of achievement in English, particularly if the instructors are not (and they are not always) English language instructors, but they are involved in teaching something other than English aerosciences, but they are teaching through English. Following Davies (1965), we used tutors' estimates as an external criteria, which might be unreliable but should have good validity. He commented as follows:

> 'their estimates might be unreliable but they should have the virtue of good validity since they are obviously concerned with the degree of English communication their students possess for their physics, medicine, history and so on'
>
> (Davies 1965, p.58)

So far as the second external criterion is concerned, (the cadets' examination results) or end of term grades. It is assumed that the greater the command of English a student possesses the higher the chances of success in his speciality. This certainly is the view of those teachers who teach their courses in English. Rea (1985) commented on this matter as follows:

> 'For several decades now, there has been an interest in correlates of university performance and, in particular, in the relationship of English language proficiency with academic success for students for whom all or part of their higher education is conducted through the medium of English'
>
> (Rea 1985, p.1)

Problems with the use of the second criterion are discussed in

chapter seven which concerns itself with the validation study of the newly designed test Battery.

4.5 Problems in proficiency testing

A proficiency test constructor meets with two main problems, (a) theroretical problems and (b) practical problems (Davies 1965). The theoretical problems are to do with defining the notion of "proficiency". He has first to decide on the language test theory(ies) that he would work within (this has been discussed earlier see (4.3)).

So far as the practical problems are concerned they can be divided into two:

(a) problems to do with sampling

(b) problems to do with administration

Sampling for a proficiency test poses a problem for the reason that there is no clearly defined syllabus to draw from. In addition the universe for proficiency testing is so vast, and a full linguistic validity would be extremely difficult to achieve. Sampling for the three levels of language, i.e. lexis, phonology and grammar, is difficult. It is particularly difficult in the case of lexis (Ingram, 1963). On what basis should one select the lexical items, should they be based on frequency counts, or their functional loads? There are similar problems for phonology and grammar.

However, depending on the situation, the language proficiency tester must make his sampling as linguistically justifiable as possible. The problems due to administration are related to the purposes of using the language test, if a test was designed so as to measure ability in speaking or writing, the test constructor has to

think of the practical problems such as administering these tests as well as providing frameworks for scoring the tests. Of the practical problems there is the availability of necessary language laboratories if listening comprehension tests are to be included in the battery etc.

4.6 Test framework

In spite of these problems in proficiency test construction the test constructor has to do what he can to take "a pragmatic" approach to the construction of his test' (Davies, 1965).

However, producing a valid measure of language proficiency depends entirely on the extent to which it reflects the situation intended for i.e. if proficiency in oral expression is fundamental to follow a particular course of study (as in our situation) and yet is not tested, the test might be incomplete and therefore lacks part of its content validity, since an ability in relation to future need is not included, therefore prediction might be difficult to predict.

With regard to our proposed test, the argument may be presented as follows:

A cadet at the K.F.A.A. needs English in the first place to follow his chosen course (in this case Aerosciences as well as flying training) and take his examination; oral and written tests of reading, listening and speaking will be introduced as measures of his English in relation to his ability to follow his course and take his examination, as measured by end of term grades, and by instructors' estimates.

It was therefore decided that for the purposes of this study, the construction of sub-tests in this test battery would include

tests of oral expression, reading comprehension and listening comprehension.

4.7 Job study

Previous work (Davies, 1965; Weir, 1984) suggested that in proficiency test development there should be two areas to draw on for the test content. These are: the language at work aspect (work sample tests) and the linguistic aspect. The job study suggested that a well-designed Battery must contain some tests which are based on linguistic categories and others based on work samples of what a cadet actually has to use his English for, in his academic studies.

Even at an early stage in the E.L.T.S. development it was proposed that a test in linguistic categories (grammatical structures and vocabulary) be included in the E.L.T.S., however at a later stage the idea was abandoned. In this respect B.J. Carroll wrote:

> 'At an early stage in ELTS Development it was proposed that 50-item tests of Grammatical Structures and vocabulary be devised, based on lists of commonly-used items of each type. These items would be discrete, or independent of each other, and not consciously derived from the candidate's specified training needs. At the later stage, no prior selection of language items was made as these were to be derived from the functional analysis and dictated by the authentic texts selected to elicit the linguistic behaviour in handling such functions'.
>
> (B.J. Carroll 1983, p.96)

Following previous work cited above, the battery would include the following groups of tests: work sample reading, work sample listening, linguistic categories (grammatical structures and oral

expression).

4.8 Work sample justification

The question to be raised is: what, within the limits of the listening, reading and speaking skills, does a cadet of the K.F.A.A. have to do with his English? The answer is: he must listen to his lectures and his instructors and understand them, he must listen, understand and speak to his flying instructors and traffic controllers; he must read his textbooks and lecture notes and understand them. Thus, the work sample tests that should be included in this battery would involve comprehension of typical textbook and lecture-notes material and comprehension of typical lecture material.

In proficiency testing designed for heterogeneous groups it is always difficult to identify what is typical. The test constructor is faced with a problem as to what would be typical for these groups. In the case of test batteries such as the E.P.T.B. the constructor decided to make a rough decision on areas of work sample tests. A gross difference between scientific and non-scientific students was recognized and therefore work sample tests in the E.P.T.B. catered for this problem by taking this decision. However, in Britain, at a later stage, it was recognized that proficiency tests should be more specific so as to cater for the diversity in the students' intake in the British universities. So, in the E.L.T.S., for instance, a finer distinction among students was made so as to include work sample tests for every sub-speciality. There is for example, a test of reading comprehension for medicine students, Engineers etc., but still these tests need to be more

specific in the sense that Engineering comprises more than one sub-speciality not counting other specialities. Precise work sample for proficiency testing, as Davies (1965) rightly pointed out "may well be an impossibility". However, in this project we do not face this problem. With work sampling tests we are dealing with a homogeneous group (trainee pilots) who are all entering an identical course of training which makes use of one set of teaching materials and one set of lecture notes with one set of instructors.

It was therefore decided that the following tests be included in the work sample section of the Battery: listening comprehension 'specialised' and reading comprehension 'specialised'. Our work sample tests may well be described as achievement rather than "proficiency", but as noted earlier achievement testing is a kind of proficiency testing.

4.9 Linguistic justification

The part of the Battery drawing on the linguistic aspect should reflect some linguistic categories. As far as dialect is concerned, it should reflect standard British English, the dialect of educated native speakers of English, and the one which is most likely to be used by instructors in the classrooms (95% of the instructors are native speakers of British English); and so far as pronunciation is concerned, it should reflect received pronunciation (R.P.) the one which cadets are expected to understand. So far as 'language components' are concerned, a separate test of grammatical structures was included. Lexis was not to be tested separately since it entered most of the points if not all in the battery. Context of situation was not tested separately because it entered into the work

sample tests. The part of the battery drawing on the linguistic aspect, therefore, is grammatical structures only.

4.10 Test of English for Air Cadets (T.E.A.C.)

General characteristics

The purpose of the Test of English for Air Cadets was to measure each cadet's linguistic and communicative abilities with the emphasis on both the linguistic ability as well as communication. Five subtests were projected. Integrative tests in which the cadet had to interact with the listener (in the case of oral production tests), or a speaker (in the case of dictation and listening comprehension tests) or simply with the text in the case of (cloze) were developed. In addition, linguistic ability was also assessed through a number of discrete item tests.

4.11 Test preparation

Material for the test T.E.A.C. was written and recorded by the writer. A battery of five tests was projected. These consisted of:

Proficiency type	Language skill	Test type	No. of Items	Type of Assessment
A Linguistic	Reading	Grammatical structure	73	Objective
B Overall	Reading (work sample)	Cloze	60	Objective
C Functional	Listening (work sample)	Listening Comprehension	13	Objective
D Overall	Listening	Dictation	68 & 73	Objective
E Functional	Speaking	Interview	-	Subjective

The test of Listening Comprehension (the simulated lecture) was not possible until after a number of simulated lectures had been recorded both at the Academy, at home and at the Air Service

Training Academy in Perth, Scotland. As for the tests of Reading Comprehension, textbooks as well as lecture notes were requested from home, the object of this was to obtain realistic test material for tests (A, B, C and D). Below we will examine each subtest in terms of its content and give brief item illustrations.

4.12 Test Material: Examples of Items

Test A: Grammatical Structure

Item writing

In general a criterion-referenced approach to item writing was used. Items were constructed each measuring a specific criterion in two categories, language and aerosciences. As a basis for some linguistic categories, the writer analyzed a number of reading comprehension passages (from the students' textbooks) in an attempt to identify those grammatical points most likely to be encountered by the cadets when dealing with subject related materials, i.e. his speciality. The grammatical structure is almost an essential part of every known or unknown proficiency test. This test consisted of 73 items and was an attempt to elicit the cadet's command over some of the ordinary grammatical structures of what constitutes the English sentence.

Although an ideal test of competence (proficiency) in the grammatical structures of a language is to test all the grammatical points of that language, one would actually need thousands of items to claim complete adequacy for his sampling but for the purpose of including a Grammatical Structure as a subtest in a proficiency test battery it would be difficult if not impossible to include all the grammatical points in a test of an acceptable length.

However, our sampling of the grammatical points to be included in our test was justified as follows: first, since we are dealing with a homogeneous population, having a shared first language (all speakers of Arabic) we made use of common grammatical errors made by Arab students learning English, therefore we included a test of problems as a cross-section. Findings of a published research project conducted by Scott and Tucker (1974) were made use of. In their study, Scott and Tucker found that: Articles, Verbs, Relative Clauses and Prepositions were the most frequent errors in Arab's writing and oral production. It was found that these errors persist before and after an intensive English course was given. Although this is not the place to discuss these errors in detail and to attribute them to influences or causes, a brief account to clarify the matter might be useful.

Tense in both Arabic and English differs in the use of the auxiliary verbs: Arabic has no auxiliary or copula, so far as prepositions are concerned they represent a tremendous difficulty to all learners of English as an L2 from different language backgrounds, they are particularly difficult for learners of Arabic language mother tongue because prepositions seldom have a one to one correspondence between English and Arabic. An Arabic preposition may be translated by several English prepositions while English usage may have several Arabic translations. As for errors in articles and relative clauses, they were largely attributed to mother-tongue interference.

In a way we are not rejecting Lado's approach to language proficiency testing, for the reason that our testees have a shared

L1. Lado's approach is summarised in the following:

> 'The non-native speaker has a mother tongue, his mother tongue may differ from the target language; identify areas of differences (contrastive analysis) and test them'.

It should be noted that the 'contrastive analysis' approach was totally rejected by the constructors of the T.O.E.F.L. and partially by Davies in the E.P.T.B. The reason is simply that these tests were designed for candidates of hundreds of L1 backgrounds and it would have been difficult or impossible to accommodate all differences in any one test designed for heterogeneous groups.

In addition a linguistic analysis of the code feature of typical text and lecture materials in Aerosciences was carried out. It revealed that the grammatical points occurring most frequently in these texts which may or may not cause difficulties for students of Aerosciences were:

a. passives, present simple tense
b. conditional and relative clauses
c. logical connectives
d. modifiers, qualifiers
e. comparisons
f. models

A number of items were constructed to test these grammatical points

4.12.1 Content of the test:

The particular aspects of grammatical structure which were included in the composition of the test items were:

- Tenses including passives and word order, "statements and questions" 25 items

- Prepositions — 12 items
- Conditional and relative clauses — 13 items
- models and auxiliaries) questions and statements — 6 items
- Logical connectives, comparisons, modifiers, articles) questions and statements — 17 items

The full text of Test A is provided in the Appendix. Here are some examples:

4.12.2 Examples

The testees were asked to indicate by marking a letter on their answer sheet (a separate answer sheet was provided). It has been devised by the Defense Language Institute and is being used by the K.F.A.A., a copy is provided in the Appendix.

Which of the four choices was the most 'correct', likely to be chosen by a native English speaker.

E.g. 1. "_____ are you flying solo?"
"next month"

a. why
b. when
c. where
d. how

E.g. 2. Airframes are normally made _____ aluminium.

a. of
b. by
c. off
d. for

E.g. 3. Water _____ of hydrogen and oxygen.

a. consisting
b. has consisted
c. is consisted
d. consists

The correct answers are as indicated.

4.12.3 Instructions for test A

Instructions were written on each booklet. They read as follows:

> This is a test of how well you can recognize correct English grammar. Each item in this test is a short sentence or a short conversation with one or more words missing. Choose the ONE answer that is grammatically correct.

The time of 30 minutes was allowed for test A although it was anticipated that only the slow cadets would need so long.

4.13 The dictation

The purpose of the dictation test was to test the cadet's ability to listen, understand and write information dictated to him. The test in a way resembles the note-taking task in a classroom situation with the exception that in the dictation test the cadet is required to write information as he hears it (i.e. in full). The texts to be dictated were drawn from the type of text the cadet is going to deal with at the academy.

The texts chosen were taken from "Strikemaster IYK80A. Student Study Guide, Volume I. The book is published by the R.S.A.F. and given free to every cadet. The book is to assist student pilots in their flying training on the Strikemaster training, at the King Faisal Air Academy (K.F.A.A.).

The first text is entitled The Aircraft. It consists of 68 items. The second is entitled Your instructor and consists of 73 items. Both texts are taken from the introduction of the textbook and did not contain highly specialised material.

4.13.1 Method of testing

The dictation was recorded on tape by a native English speaker. The following instructions were included on the tape: instructions for punctuations although included on tape, they had to be explained on the blackboard as well:

> This is a dictation. You will be given two paragraphs to write down. The first time you hear each paragraph, just listen. The second time there will be pauses for you to write down what you hear. The third time you will be given marks of punctuation. When you hear 'full stop' do not write down f-u-l-l - s-t-o-p but put down a dot.

The dictation test is one part of the Test of English for Air Cadets introduced in the battery. It was not merely a text to be read with pauses but candidates were required to listen, understand, and write information at certain points.

Similar to the "structured interview" in this test battery, the dictation had three phases. The first was the introductory in which the speaker read the instructions and introduced the texts as a whole, so that his voice became familiar to the candidates and the content of the texts become known to them as well. Candidates at this stage were required to listen only. In the second phase, the speaker indicated by giving new instructions that he is going to dictate the texts one by one. He asked the students to be ready. Between pauses, students were required to write down what they had

heard as accurately as possible.

At the third and the final phase, the speaker gave new instructions to the candidates. At this stage, candidates were required to write down marks of punctuation. The speaker read each text for the third time during which he was asking candidates to write down marks of punctuation where appropriate. At the end of this phase, the speaker told the students that this was the end of the test.

Because the dictation was to be given to several groups and at different times, it was essential to make sure that the presentation was the same for all. The texts, therefore, had to be recorded and the length of the pauses had to be decided. The length of the pauses were determined by dictating one of the texts to an average non-native learner of English and observing the time he took to write down each section of the text. The minimum length of each section was a group of two words and the maximum was six. The length of the pauses varied between three and sixteen seconds. The full texts are included in the Appendix.

4.13.2 Method of scoring

The dictations were scored by counting the number of correct words appearing in the sequence dictated. As long as no phonological rule of English had been violated, misspelled words were considered correct. For instance, "personel" was accepted for "personnel" but "bilot" was not accepted for "pilot" or "nefer" for "never". In addition to clear errors in spelling and phonology, grammatical words 'go' for 'went' or choice of wording 'were' for 'where' were also considered as errors, therefore no allowances were given.

Texts were scored separately, two scores were recorded for each testee.

Time: 15 minutes were required for completion of the dictation test.

4.14 The cloze (the reading test)

The purpose of the reading (cloze) test was to gain some indication with regard to candidates' ability in reading comprehension. Completion of the cloze tests required the candidates to interact with the text.

Several points were raised with regard to the content of the cloze test. These were:

The number of passages to be included in the test

The rate of deletion to be adopted

The topics

The number of deleted items

4.14.1 Four passages were chosen for the reading test. Two were based on material taken from the students' own text books and lecture notes. These materials represent the only source of reading material the students have, and they are not referred to references other than their textbooks and lecture notes. Two texts were taken from the students' text books and lecture notes produced by their instructors (native-speakers of English). One came from the subject of 'Meteorology' and the other came from the subject of 'Airmanship'. The other two were taken from a "General Aviation English Course" published by the "Air Service Training Academy" in Perth. Every passage contained 15 deletions making a total of 60 deleted items.

4.14.2 Criteria for selecting the passages

The texts were selected on the basis of the following criteria:

(a) The passages were not accompanied with any illustration, diagrams or mathematical formula in order that no information other than that provided by the reading texts could be used in providing the missing words.

(b) Each text had to be a self-contained unit with respect to content. It did not refer to remarks either before it or after it.

(c) For the purpose of comparison, two texts were chosen to represent authentic materials and the other two were simplified aerosciences materials. The selection of two simplified and two unsimplified texts was deliberate for the following reason: it was assumed that authentic materials may prove more difficult than unsimplified material in the sense that authentic texts may include some lexical items related to the subject matter itself which cadets may or may not have encountered. For a pilot test, I think, it is desirable to include variety of texts in order to test their difficulty for one does not want to construct a test which proves to be difficult and consequently causes students to fail. Authentic materials are highly technical in content. It includes lexical items which may prove problematic therefore makes comprehension more difficult.

4.14.3 The deletion rate

In the existing literature on cloze testing, the most commonly used deletion rate is the systematic deletion type where every n'th

word is deleted, n'th has been defined as every 5th, 6th, 7th or 8th word. However, a consistent deletion rate was adopted for all the four texts since it was expected that some cadets might not complete the four passages. A systematic type of deletion was used resulting in both syntactic items as well as lexical items being deleted. Every 5th word was, therefore, deleted. The first and the last sentences for each text were left intact to provide lead-in and lead-out context.

4.14.4 The characteristics of each of the four texts are given below. The original texts appear in the Appendix.

Text A

Subject:	Airmanship
Style:	Academic discussion
Type:	Authentic
Source:	Airmanship (student textbook)
Length:	Instructions, plus 118 words
Number of Deletions:	15 items

Text B

Subject:	Meteorology
Style:	Academic discussion, descriptive
Type:	Authentic
Source:	Students textbook
Length:	Instructions, plus 115 words
Number of Deletions:	15 items

Text C

Subject:	Aviation English
Style:	Descriptive
Type:	Simplified aviation material
Source:	A general Aviation English Course, published by the Air Service Training (A.S.T.) Perth, Scotland
Length:	Instructions, plus 106 words
Number of Deletions:	15 items

Text D

Subject:	Aviation English
Style:	Narrative
Type:	Simplified aviation material
Source:	A general Aviation English Course, published by the Air Service Training (A.S.T.) Perth, Scotland
Length:	Instructions, plus 111 words
Number of Deletions:	15 items

4.14.5 Instructions

Written as well as verbal instructions were given to the subjects. The written instructions were provided at the top of each text. An illustrative example was also provided. The instruction read as follows:

> 'Read the whole passage first. Then go back and fill in the blanks with the one word you think is missing. The following is an example:

e.g. I was born on Tuesday the 25th of June.

Testers were also requested to give verbal instructions to students and if necessary they were also asked to write an illustrative example on the blackboard. Students were also told that contractions count as one word, e.g. don't.

4.14.6 Scoring method

For the purpose of this study, it was decided to adopt the acceptable word method; responses were counted as correct only if they fit all of the surrounding context. A contextually appropriate word was defined as what was considered to be grammatically correct and contextually appropriate by a native speaker.

4.15 The interview (the speaking test)

The aim of the interview was to assess the subject's ability in communicating orally in English allowing him to speak at some length

on several different topics. The role of the interviewer was to incite and stimulate the testee to speak. The interviewer's job was also to steer the interview, giving the subject the opportunity to speak freely so that the assessor would form an opinion (give a score) of the interviewee's performance in the oral production of English. The general idea was to give the subject control in so far as the oral exchange was concerned. Since it was anticipated that proficiency of students would vary it was decided that 'a structured interview' was deemed necessary and it was therefore planned to conduct the interview in stages.

4.15.1 The first stage would be an introductory one. The dominant party at this stage was the interlocutor who would, in a minute, explain the purpose of the interview and put the candidate at ease. It was important at this stage to indicate to students (particulary those who have just passed the E.C.L. test), that the main purpose of this interview and indeed of the other subtests is research. It was also stressed that the results would have no effect whatsoever on the student and would be kept secret.

At this stage also the interlocutor would make sure that all other personal details of the interviewee are checked and completed, i.e. their names, class and service number were documented. The purpose was to be able to trace the candidate's other scores in either the E.C.L., end of term exams, his scores in the other subtests and on the English ability rating forms completed by his instructor.

4.15.2 The second stage of the interview mainly consisted of 'wh' questions. It was anticipated that at this stage the subjects may not experience any difficulty. The questions were mostly formulated

on the subjects' personal background. Questions at this stage were aimed at the following topics:

- Home province
- Education
- Father's occupation
- Travel abroad, purpose
- Reasons for joining the Academy
- Hobbies

However, this was no way a thorough list and since oral communication is unpredictable, the interviewer had to make use of the subjects' answers to create new situations and therefore new formulated unanticipated questions. This stage formed the basis for the third stage.

4.15.3 The third stage was designed in such a way that the interviewee would have control over the oral exchange. It formed the narrative stage in which the candidates spoke for a good length of time. It was planned that subjects should not be interrupted unless it was felt that they needed some help or the discourse was becoming repetitive. Questions had a descriptive nature at this stage. The subjects were asked to describe some events they had encountered. They were asked about the following:

- A description of a typical day in the Academy
- A comparison between life in the barracks and life at home
- A description of the interviewee's home town
- A description of the interviewee's home town

4.15.4 In the final phase of the speaking test the interviewee was asked to put any question or questions he might have to the

interviewer. At this stage the interviewee took complete control over the interview. This stage had a number of aims. First it gave the assessor a chance to note the subject's ability to form questions in English. Second, it enabled the testee to inquire about anything he had in mind with regard to the current exercise or otherwise. If the subject asked questions, the interlocutor would therefore respond as fully as possible and would have a chance to note the extent to which the subject was able to interrupt the discourse and make it a near life situation communication.

The interlocutor then had to bring the interview to an end by making an appropriate gesture, making sure that he thanked every testee he interviewed.

4.15.5 Scoring the interview

The main problem with productive tests such as the interview as a speaking test, is the framework or the criterion by which the performance is to be measured. The interview, as illustrated above, is introduced to assess the subject's overall communicative ability. Subjects were being evaluated on the basis of several factors. These were: the content of their responses, the accuracy of their language, their understanding and their fluency in an overall communicative situation. Their speech is assessed within these factors mentioned above.

4.15.6 Several frameworks for assessing oral interviews have been proposed. On the other hand, there is the analytical approach where certain segments of the performance are isolated and assessed individually. The F.S.I. oral interview uses this approach as scoring procedures. However several questions arise with this

approach.

Firstly, what are the features to be isolated and therefore tested?

However, if the features become too many, it would, therefore, be very difficult for the assessor to give an accurate assessment particularly if he is the interviewer. Secondly, there is the question of weighting. Having isolated the features. Are they to be treated equally in terms of scoring or what? And the final question is with the scale. How many scale points should there be? This is to do of course with the feasibility factor. However, if the scale becomes too long it may prove difficult for the assessor to give an accurate judgement particularly if he has too many features to observe.

However, in the context of the oral proficiency interview, the following features could be assessed individually:

(a) accuracy of syntax

(b) accuracy of phonological features (pronunciation)

(c) the range and the appropriacy of the lexis used

(d) fluency of the subject

(e) extent of subject's comprehension as well as the interviewer

(f) effectiveness of comunication

A different approach is to adopt an overall performance scale with a definition to each point in the scale describing the type of. performance the subject produces. There were two alternatives either to adopt a scale with an even number of points or a scale with an odd number of points. The advantage of an odd number is

that it contained a mid point, denoting satisfactory or average performance. With three points for below and three points for above, for the purpose of this study, a scale of 1 to 7 points seemed appropriate.

4.15.7 The definitions of the points are given below:

Level 7. Native speaker ability

Level 6. Near-native speaker ability and fluency with occasional evidence of non-native speaker pronunciation features

Level 5. Full communication established, but with occasional minor faults of English usage and pronunciation

Level 4. Adequate communication, sometimes impaired by a number of minor faults of English usage with occasional evidence of hesitations

Level 3. Communication exists, but dominated by hesitations and frequent errors in grammar/ lexical choice and pronunciation

Level 2. Communication is limited to answering 'wh' questions, with serious errors in grammar and pronunciation

Level 1. No communication at all

The full content of the interview as well as the scale points are given in the Appendix. A full transcription of the content of three different interviews is also given in the Appendix.

4.16 Listening Comprehension Test

Although an ideal work sample test of listening comprehension for the cadets would be a live recording of a typical lecture presentation with all the background of lecturing, i.e. discussion, questions to be asked by students, clicking of chalk, referral to a visual aid and so on, recording of such materials was avoided on three grounds, these are: Firstly, all Aerosciences lectures are

normally visual (often real) related discourse. Instructors' presentations in almost every topic on aviation sciences is accompanied by the use of a visual aid. Such recording for the purpose of a test on listening comprehension tend not to be suitable, for the lecture would include many references to the 'visual aid' and therefore referral to it is part of comprehension. However, if the lecture was to be recorded the test would have been incomplete. Secondly, with the exception of introductory lectures, typical lecture materials tend not to be suitable to extract a short passage from for testing, either because the lecture makes references to previous lectures, therefore, comprehension is connected with what has been taught, i.e. the lecture might not be self-contained and would not be adequate for testing purposes. In addition a live lecture might also be highly technical i.e. difficult in content, so it would not be fair to include it in a language test of this nature. Thirdly, the tester should also be concerned with practical matters such as clarity and audiability of the test presentation. However, in a live lecture, unless the lecturer stands still (and this does not always happen) adequacy of the presentation in terms of the factors mentioned above might be difficult to achieve.

In a listening comprehension test, the tester's primary consideration is to find out whether testees have or do not have the skill, in other words, he is concerned with whether listening comprehension as a skill is present or absent on the part of the testees.

However, for the purpose of this study, a number of simulated

lectures were recorded, but only one was found to be suitable in terms of content and delivery. The others were found to be either very difficult in content or the rate of presentation was not suitable. This lecture was an introductory lecture on Meteorology. It was given by John Venables, the Meteorology instructor at the Academy, to whom the writer is very grateful. The cadets had the advantage of being familiar with his voice and presentation.

4.16.1 Method

The lecture was heard but not seen. Testees were allowed to take notes during the presentation. After the lecture was presented the students were given fifteem minutes to answer thirteen questions. The questions were in a multiple choice format, each with four alternatives. They were asked to tick the one answer they thought was correct. All the questions were based on the content of the lecture so full understanding and concentration on the part of the testees was required.

Here is one example of the questions asked:

A jet stream is:

(a) an easterly wind

(b) A squally wind

(c) a wind blowing at 100 knots or more

(d) the wind at cruise level

The correct answer is as indicated.

Transcription of the lecture and the full test is provided in the Appendix.

4.17 Testing methods of the T.E.A.C.

With the exception of the speaking test (the oral interview, although a scoring framework was worked out to make assessment of speaking reliable) all other sub-tests in the battery were designed in such a way that they could be scored objectively.

The listening test (lecturette) and the Grammatical Structure test contained multiple choice type questions, each with four alternatives. The reading test (the cloze) was objective but not multiple choice. It was objective in the sense that only one type of answer is counted as correct and that is if it is contextually and grammatically appropriate. The listening test (the dictation) was also objective in the sense that each word is given one mark as long as the word appears in the original sequence of the passage and does not violate an English phonological rule (see 4.13.2 for details).

For the multiple choice questions (i.e. the Grammatical Structure test and the listening test (the lecturette), a separate answer sheet for each was used. This answer sheet was developed by the Defense Language Institute, Texas, U.S.A. and is used by the K.F.A.A. for conducting the E.C.L. test. As far as the cloze test was concerned, the cadets were instructed to write on the same test sheets. In the case of dictation, students were provided with ordinary lined papers to write on.

4.17.1 The listening tests

The listening tests (the dictation and the listening comprehension test, the "lecturette") needed more preparation. They had to be efficiently recorded. Through the good co-operation of my

colleagues here and at home these tests were recorded. The dictation test was recorded by a colleague in the Applied Linguistics Department, University of Edinburgh, a native speaker of English, who spoke with an R.P. accent. The lecturette was recorded by another colleague (the instructor of Meteorology) at K.F.A.A., who was also a native speaker of English and spoke with an R.P. accent too.

4.17.2 Test scoring

As was mentioned in the previous paragraphs all sub-tests in the battery were objective, and it therefore follows that scoring was objective too. All the tests were scored by the writer with the exception of the speaking tests, which were conducted and scored by native English speakers. One mark was awarded for each correct answer and the final score was the total of correct answers.

4.17.3 Number of items

The number of items appearing in the total Battery was as follows:

- Grammatical structure 73
- Dictation 68 & 73 (two texts)
- Cloze 60
- Listening Comprehension (lecturette) 13

In addition to the speaking test.

4.18 Formulation of hypotheses relating to the T.E.A.C

Apart from the major hypothesis which has already been set up and verified which relate to the testing system currently in use by the Academy (the A.L.C.P.T. or the E.C.L.) a number of hypotheses which relate to the newly designed test are also set up.

These hypotheses relate to the regular research on language proficiency testing as well as current debate on language proficiency.

The work of Davies 1965, Alderson 1978, Moller 1982, Weir 1984, for instance, began with certain assumptions and hypotheses.

The present work has also started with certain hypotheses. These hyphotheses relate to:

4.18.1 Hypothesis one

In view of the claims in the literature (Oller 1979 and 1983) that proficiency in language is unitary, that there is only one factor underlying performances on language tests, the null hypothesis was set up: that language proficiency is unitary and that there is a general factor underlying performance on language tests aimed at assessing proficiency.

4.18.2 Hypothesis two

Since the validity of a language test is considered of the highest importance the null hypothesis was set up: that there was no positive relation between English proficiency as measured by the T.E.A.C. and either instructors' estimates or academic success in aerosciences.

4.18.3 Hypothesis three

Since two distinct samples of cadets were selected for testing purposes, the null hypothesis was set up: that there is no significant difference between level one and level two in terms of language ability.

CHAPTER FIVE

TEST OF ENGLISH FOR AIR CADETS, T.E.A.C

FIELD WORK AND PILOT EXPERIMENTS

5.1 The journey to Saudi Arabia for field study

The writer consulted with his supervisor regarding the name to be given to the test battery and the characteristics of the sample to whom the test was to be given (see 5.8 below). For the name, he suggested "Test of English Air Cadets", T.E.A.C. for short. The test Battery, therefore, was labelled T.E.A.C. I then requested permission to leave Edinburgh for Saudi Arabia to apply the test Battery and to distribute the Academic English skills questionnaire, and collect the relevant data for the research project such as the instructors' estimates of cadets' language proficiency in accordance with an English ability rating scale originally devised by Moller (1982), but expanded and modified by the writer. The necessary administrative procedures were made, and by the middle of June 1985, I had left the U.K. for Saudi Arabia. The period mid-May to mid-June 1985 corresponded with the holy month of Ramadan, during which Muslims fast from dawn to sunset. In Saudi Arabia, Universities, Colleges, Schools and all other teaching establishments close during the month of Ramadan, either the whole month or part of it. The K.F.A.A. in which the test Battery was to be applied was no exception, it was closed too, and would not reopen until the end of June.

5.2 At the Academy

Facilities and difficulties

The Academy, in the first place, is a military institution. Like

any other establishment anywhere, security measures and rules are literally implemented. Any research projects, changes in academic programmes, distribution of questionnaires and so on have to be brought to the attention of the Commandant and the Officer commanding of the relevant Wings.

Although the Commandant and the officer commanding Academic Wing were both fully aware of my research project (after all, they nominated me to conduct this study), I had to ask their permission to apply the tests and distribute the questionnaires to the Academy's instructors. Permission was granted. The second step was to arrange for a meeting with the English Language and Aerosciences instructors to brief them on the aims of the research project, to orient them of the content of the test Battery and discuss with them matters related to the administration of the test Battery.

5.3 English Language Instructors meeting

Before tests were duplicated, it was necessary that a seminar be held so as to orient all instructors of advanced English classes (cadets who have spent six months or more in the language programme and have had their language proficiency assessed by the E.C.L. but have not qualified yet, i.e. scored 75% or above). Instructors of these cadets who have had their proficiency assessed by the E.C.L. and scored 75% and above, but were waiting to begin their training in Aerosciences, or have just, at the time, begun their training in Aerosciences (i.e. have spent three weeks or less), were also invited.

The meeting was held in the beginning of July 1985 and was attended by the following instructors:

Mr. J. Adams	English Language Instructor
Dr. N. Davies	English Language Instructor
Mr. J. Hughes	A former pilot and a senior English Instructor.
Mr. R. Kirby	Senior Language Instructor
Mr. W. Knight	English Language Instructor
Mr. P. Ryan	English Language Instructor
Mr. P. Stalker	Senior English Instructor
Mr. C. Suttie	English Language Instructor
Mr. P. Topliss	English Language Instructor
Mr. R. Triggs	English Language Instructor

Thanks should be expressed to them for their cooperation.

Firstly, the meeting was a good opportunity to hear and document their reactions to the new developed testing instrument (i.e. to gain information on the test face validity). Their reactions will be discussed separately in this chapter. It was also a chance to hear their comments on the technical content of the "grammatical structure test items" and to arrange with them procedures for applying the tests.

After going through the test battery items, Mr. Hughes (a former pilot and an English language instructor) drew my attention to items number 13 and 68 of the grammatical structure test. In the original format, item 13 read as follows:

(13) Aircraft frames are normally made light steel.

(a) of (b) by (c) off (d) for

He pointed to the technical content of the item as being erroneous. He said that "Aircraft frames" should be replaced with 'Airframes' and 'light steel' should also be substituted with 'aluminium'; he commented that an Aircraft would never be able to take off, if its 'Airframes' were made of steel whether light or heavy.

So far as item number 68 was concerned, the type of 'Cessna' included in the item should be 'Cessna 172' and not 167. The latter

number is reserved for the 'Strikemaster' type of aircraft The content of the items were corrected and thanks were expressed to Mr. Hughes for his observations. The grammatical points being tested in these two items were retained.

Since it was planned to give the test to all advanced English learners as illustrated in the previous paragraphs, the total amounted to "103" cadets. It was therefore necessary to seek help in administering these tests. All the above mentioned instructors agreed to help me in conducting these tests and thanks once again should be expressed to them.

5.4 Printing facilities

By good luck, a more advanced electronic printing facility had been installed in the Academy. A lot of time and effort was saved in duplicating the test Battery. For security reasons, only one hundred and ten copies for each test were duplicated. Sufficient copies were also duplicated for the Academic English skills questionnaire and the English Ability rating forms. Three different colours were chosen to duplicate the tests, the rating forms and the questionnaire. While the white colour was reserved for the test Battery, pink was chosen for the English Ability Rating, and light blue was chosen for the Academic English questionnaire. With the help of some non-commissioned officers, the materials were printed and stapled together. They are to be sincerely thanked for giving of their own time and effort.

5.5 The aims of trialling the T.E.A.C.

The aims of the pre-pilot experiment were:

(a) To gain experience in applying the integrative type of tests

included in the Battery; the cloze, the dictation and the listening comprehension test. Cadets were familiar with the multiple choice procedures "discrete-item test formats" but not with the above mentioned tests.

(b) To gain information about suitable duration for each subtest and about ease of administration and scoring.

(c) To gain information about the suitability of the length of pauses between 'chunks' in the dictation test.

(d) To gain experience about conducting the speaking test and assess raters' reliability of scoring the interview.

(e) To gain some normative data particularly on test inter-correlations, although conclusions reached at this stage would only be viewed as tentative.

5.6 Test administration

As in any other educational institute, cadets were strictly following a pre-planned programme. It was even more strict in military institutions. From six o'clock in the morning till 12 o' clock in the afternoon, cadets were either taking their academic classes or flying. In the afternoon, they have a different programme, marching and sport classes occupy the whole of the afternoon. It was therefore necessary to give the tests in the morning and a disruption of the programme seemed inevitable. Arrangements with the Officer Commanding Academic Wing had to be made to conduct these tests in the morning, during the ordinary class periods. The period 7.30 a.m. to 9.30 a.m. seemed a suitable time to give these tests. Arrangements for every test had to be made separately with each class instructor. With the exception of

the oral interview, all tests were conducted by me and my colleagues whose names were mentioned previously. The availability of six laboratories with a capacity for 120 cadets helped make the job relatively easy.

Before we give any specific details about the pilot study and the characteristics of the sample it is worth at this point documenting and discussing the reactions of the English languge instructors in the Academy (all native speakers of British English) to the test Battery.

5.7 Reactions to T.E.A.C.

The English language instructors at the Academy (all native speakers of English) whether they had participated in the actual administration of the tests or not, were invited to comment on the test battery. The choice of the "language instructors" was deliberate on two grounds. Firstly, they were themselves professionally involved in the process of language teaching and testing, so any comments coming from professionals would be technical in content. Secondly, they may well be the users of the test and would help at a later stage in the actual production of the language teaching materials in accordance with the findings and the recommendations of this study. Comments were provided, often detailed and always helpful and critical. It was not totally unexpected, because they came from people who were directly involved in the language teaching profession at the Academy, some for a decade. Their comments and reactions are going to be illustrated and discussed below. The test in general was well appreciated for its obvious face and content validity (skills tested). There was

approval for the notion underlying the T.E.A.C. One instructor summed up his view as follows: "The test approaches the language proficiency level necessary for flying training and associated Aeroscience subjects".

Another commented as follows: "the speaking and listening sections would broaden the range of language skills needed by the cadets to follow their speciality". A third (a very experienced instructor indeed, and a former head of the English department) in a lengthy comment on the Academic English skills questionnaire explained the long felt need for such a test and that it should have a good backwash effect. It is worth quoting his comments in full: "If you want the cadets to be able to understand easily explanations in English, to respond quickly and appropriately to instructors, to speak fluently, to formulate questions easily and so on, you must make it worth their while to acquire these skills by testing each skill and by insisting that the desired level of proficiency in each area is achieved before the students leave full time English Language Training (ELT). Therefore I would say that the first thing to be done is to redesign the testing system (E.C.L.) so that the whole ELT programme becomes more specifically objective-oriented". This valuable comment came from a person who lived with the problem himself. He was often asked to re-train students after leaving the language programme. It was and still is often the case that cadets were sent back to us in the English Department to be given some remedial English although they were considered proficient according to the E.C.L. test results they gained.

His comment also brought to the surface the lack of content

validity the present testing and teaching programmes possess in terms of language skills taught and tested. He was particulary commenting on the E.C.L. testing system:

> 'The method of testing in ELT, which determines the direction in which both students and instructors will expend their main effort does not motivate the students to acquire useful language skills'.

Nothing will be said here about the validity of the present teaching/testing programmes, since this was discussed in the introductory chapter and chapter two.

Below, we will examine each sub-test in turn with regard to instructors' reactions.

5.7.1 The Grammatical structure test

"Very useful", "well constructed", "a good E.C.L. type test", "it approximates with the E.C.L.". These were typical comments. One critic felt that it breaks no fresh ground, although he praised the test; his comments were "A good, ECL style test, but it breaks no fresh ground". However, the instructor was informed of the rationale behind the selection of the grammatical items (see 4.12.1 and 2 for details). Another critic felt that the stimuli/responses in this test are shorter and more straighforward than might be expected". However, in comparison with other proficiency test batteries stimuli/response in this test do approximate with its counterparts in other measures.

5.7.2 The Oral Test (the interview)

"An excellent idea", "effective measure". These were typical comments. With this test, instructors were concerned with "rater reliability". One commented: "An effective measure of general

proficiency but it would be difficult to achieve the necessary degree of standardization without a team of interviewers working in close cooperation". Another commented: "An excellent idea, but impractical timewise, except by a specialist interviewer".

These and similar comments were appreciated. One certainly has to do whatever possible to make scoring reliable, but should not for practical reasons omit tests of importance to his situation. It is worth quoting Davies (1965) in full: "that there is no reason why one sole technique (e.g. objective testing has to be used to the exclusion of all other techniques" (Davies 1965, p.103).

5.7.3 The Cloze Test (the reading)

It was not at all unexpected that this technique in particular faced a number of critical comments. In general comments ranged from about the right level of difficulty to difficult. But the most obvious concern was with a central issue of cloze testing:

What is it supposed to test?

One critic commented: "I am not sure if its testing structural/ technical vocabulary or technical knowledge". Another, "The cloze tests are as much a test of subject specific vocabulary as general reading comprehension ability".

It is quite obvious that the two preceding comments express a concern regarding the cloze test in so far as what it is actually testing. Although this issue has been thoroughly discussed in the review of the literature section, it is worth repeating it here very briefly.

A testing method originally used as a measure of texts readability then as a measure of reading comprehension, but recently

it has been looked at as a possible measure of second language proficiency. However, it was introduceed in this study as a measure of reading comprehension. But, perhaps an internal analysis of the deleted items (of the texts used in this study) helps identify the nature of the test in some detail.

Two terminologies will be used to classify the nature of the deleted items. These are: functional items and content or lexical items.

Text A

Content items 6
Functional items, 9 of which:
3 prepositions
2 connectives
4 articles (definite/indefinite)

Text B

Content items 6
Functional items, 9 of which:
3 prepositions
3 connectives
2 definite articles
1 qualifier

Text C

Content items 7
Functional items, 8 of which:
2 prepositions
2 articles
2 connectives
2 pronouns

Text D

Content items 2
Functional items, 13 of which:
3 articles
5 prepositions
3 auxiliaries
1 connective word
1 pronoun

Briefly, the analysis reveals that it is as much a test of grammatical structure as lexical items and a test of context as well as comprehension. A full understanding of the grammar of the English language is needed to be able to restore the missing items.

Some additional comments were made to raise the issue of "the implications" of this study and it brings to the discussion the notion of progress and achievement tests as well, although a full chapter is going to be devoted to this issue, i.e. the implications of this study. It is worth quoting the full comment on the cloze format, as was written to the writer by one instructor:

> 'Given its 4-test format, I cannot understand why it was not conceived as a 4-step stairway up to Aeroscience level. This would furnish a neat assessment of English department needs to prepare cadets for their language future. Such testing should/could begin with sentences for completion at the outset, and build up via brief paragraphs in the intermediate phase to continuous text at advanced level'.

The issue brought up in the previous paragraph is as much a pedagogical issue as a testing issue. And both are interrelated. As Valette (1977), Lado (1961), Davies (1968), rightly point out, testing is a topic of concern to language teachers, and what is actually to be tested (in most cases) constitute what students have been learning. It is worth quoting Davies (1968) here in full:

> 'Language testing involves both linguistics and psychology because it is concerned with language and with learning'.

Looking back at the instructor's observation, we can infer from it that both teaching and testing are closely related. Proficiency testing in most cases directs teaching, i.e. sets out to evaluate language programes. While the instructor agrees with the writer of

the 'proficiency level' introduced in this sub-test, i.e. the "cloze" "... a 4-step stairway up to Aeroscience level ..." he would like to see his students in the English department trained to do the cloze procedures in a systematic way. His obvious concern and indeed the concern of all other instructors is with "achievement" and "progress tests" within the English language programme offered in the K.F.A.A. However, as noted earlier, some cadets join the Academy with a reasonable proficiency in the language, and they may not need to follow the language programme. The test, therefore, would serve as both a "placement" as well as a "proficiency" testing instrument. Placing cadets within the language programme, and a proficiency in the sense that it has been designed to evaluate required abilities in relation to future needs, progress tests were actually left for the individual instructor to construct. The progress test was viewed as a teaching device. Its objective is to reinforce what has been taught. It also allows the subject to demonstrate what he has or what he has not learnt. Our test was aiming towards proficiency in language abilities related to expected behaviour.

However, achievement and progress tests within "the (proposed) English language programme" (see Chapter Two) for the cadets of the K.F.A.A. will be a separate future project, and tasks and pedagogical exercises given to cadets should in fact encourage relevance teaching within the new testing programme. In this way we will ensure that teaching will be objective oriented.

5.7.4 The listening and the dictation tests

These tests were in general well appreciated by Aerosciences instructors. The listening test (lecturette) was particularly praised for its obvious face validity. Typical comments were: "The listening sections would broaden the range of language skills tested if incorporated with the ALC.

5.8 Characteristics of the Sample

The sample contained 103 (the total number conformed with the characteristics laid down by the criteria for selection) cadets with an average age of 20 years. ALL of them were holders of the National Secondary School Certificate (NSSC). They all learnt English as a compulsory subject at the intermediate and secondary schools for six academic years with an average of 3 periods per week. They represented a mixture of two branches of the secondary school education. These were scientific and literary departments. Some were following an intensive English course in the Academy and had spent six months or more in the English language programme. Others had completed the language programme and were waiting to begin or had just begun their Aerosciences training. The sample was chosen on the basis of the following criteria:

(a) Cadets who had had their English language Proficiency assessed by the E.C.L. and were still following the language programme at the Academy (they had scored below 75%, but represented advanced classes).

(b) Cadets who had had their English language proficiency assessed by the E.C.L. scored 75% or above but were still waiting to begin or had just begun their training (spent less than a month).

Age: The ages of the cadets ranged from 18 to 22, with a mean of 20 years.

5.9 Trialling programme (pilot study)

Arrangements were made to try the tests on a number of cadets in accordance with the characteristics laid down earlier (see 5.8). The sample comprised cadets whose proficiency in English had been assessed by the E.C.L. and had or had not achieved the required passing mark on the E.C.L. (75%). The sample consisted of the following levels:

- Cadets whose proficiency in English had been assessed by the E.C.L. and were still following the English language course, but represented advanced English learners
- Cadets whose proficiency in English had been assessed by the E.C.L., achieved 75% and were still waiting to begin aeroscience
- Cadets whose proficiency in English had been assessed by the E.C.L., achieved 75% and had just begun their aerosciences training.

Dates, times, lecture hall and laboratories were arranged for the testing programme on five separate days in the English Department. For those cadets who started their aerosciences programme, extra coaching classes were arranged for them since it was inevitable that they would miss some classes. Their instructors were very willing to give them every assistance. All tests were administered individually.

Trialling of the T.E.A.C.

5.9.1 The Oral Interview

Although a "proficient non-native instructor" of a language may

well be able to conduct and make accurate assessments of oral expression tests, provided that he is given a guide, i.e. a scoring framework to follow, this was avoided on three grounds. Firstly, the availability of native English instructors. Second, a native speaker of English as an interlocutor provides an advantage, i.e. he makes communication more realistic. Here, a foreign student is trying to communicate with a native speaker. Third, a native speaker is more able to incite the interviewee and above all can make a more precise judgement. The interview tests were conducted by three native instructors.

5.9.1.1 <u>The speaking test (Rater reliability)</u>

The purpose of this pilot experiment was to find out to what extent had experienced EFL native-speakers of English, assessing cadets oral expression, reached the same judgement (i.e. gave a similar score).

Assessors: Three assessors participated in this pilot study. They also conducted the interviews in the major tryouts. ALL were experienced EFL instructors. All were native speakers and have been teaching English in the Academy for 2-6 years. They all held postgraduate qualifications in Education, TESL and Linguistics.

5.9.1.2 <u>Procedures</u>

Assessors were required to rate speakers on a seven point scale (see 4.15.7). One of them conducted the interview and the other two observed the interview, but at the end each gave an independent score. The interlocutor gave a score too. Two assessors conducted the interview each for three cadets and one conducted the interview for the remaining two.

Table 5.1: Levels awarded by three raters for the oral interview test (pilot study)

Cadet	Rater 1	Rater 2	Rater 3
1	5	5	5+
2	5	5	5
3	4	4+	4
4	4	4	4
5	4/5	5	5
6	4	4	4
7	3	3	3
8	3	3	3

In view of this small sample, the interpretation of these results was straightforward. Within the limitation of this experiment, there was some evidence of raters' reliabiltiy. Each of the three raters deviated only in a single case. While rater 2 and 3 used the sign + in two cases to indicate performance of two subjects, rater 1 assigned two levels for one cadet. There might have been evidence of both performance in the particular cadet's performance. As was shown in the results above, raters who used the + sign were reluctant to give the cadet the following higher level as performance of the particular cadet did not quite conform to the description of the following level.

In conducting the interviews the four stages as described (see 4.15.1 to 4) were followed. The pilot study took place in an informal setting in the Academy. The introductory phase, together with efforts of interlocutors and the informal setting helped in making cadets relaxed and cooperative. In the context of trialling the measure the following observations were made:

(a) In general cadets seemed keen and to some extent fluent in talking about their personal background and own experience such as previous education, hometown, travelling abroad, family etc.

(b) Cadets liked the idea of having their spoken ability assessed.

(c) The fourth phase of the interview (asking question or questions by the interviewee) was not always possible. Nevertheless, if it existed it was always very interesting. Examples will be provided in the context of the major tryout analysis.

(d) The length of the interview varied. The minimum time was 10 minutes and the maximum time about 30 minutes. The most frequent time noted was 15-20 minutes.

5.9.1.3 Conclusion on the trialling of the interview

The trialling of the interview provided the interlocutors with an opportunity to practice the techniques of interviewing. Setting the cadet at ease, provoking cadets to answer questions which required long answers, interrupting the discourse when it became repetitive or when getting cadets to ask questions, finishing the interview, assessing cadets performance in accordance with the descriptions of the assessment scale were all techniques which needed some practice. The pilot study also highlighted that with practice and adherence to the stages of the interview described earlier, and also sticking to the assessment scale provided, could make assessment of speaking ability through the interview technique both possible and reliable.

5.10 The Dictation

The main purpose of the pilot study for the dictation test was to observe the extent to which the length of the pauses between the 'chunks' were suitable. However, in the light of the experiment, it was found that most pauses needed some adjustment. In most cases, pauses were found to be shorter than anticipated. Frequent complaints from the cadets were observed during the test. This was

appreciated. In marking the test samples the writer observed that some chunks were not fully written. However, the necessary alterations were made. This test was re-recorded in accordance with the findings of the first trial and trialled once again.

5.10.1 Results of trialling the dictation

The dictation was administered to the same group. The group found it an easy task. After pauses between chunks had been corrected for length, the test was administered for the second time to the same group. In general the group found it not a difficult task to perform. Individual scores were:

Table 5.2: Individual scores obtained by ten cadets on the dictation tests with their means and standard deviations (pilot study)

No.	Dictation one	Dictation two
1	61	63
2	58	62
3	53	53
4	58	59
5	52	64
6	61	67
7	54	57
8	47	52
9	43	44
10	36	43
Mean	52.3	56.4
SD	8.1	8.2

The higher mean for text two is due to the increase in number of items. Items were 73 for text two and 68 for text one.

5.11 The grammatical structure test

The grammatical structure test was administered to a small number of cadets as the first part of the Test of English for Air Cadets.

As was anticipated, cadets experienced no difficulty in the procedures since they were all very familiar with the test format, i.e. multiple choice questions. Individual scores were:

51, 50, 49, 48, 47, 45, 43, 38, 37, 35

Mean = 44.3

The grammatical structure results will be examined in some detail in the context of the internal analysis of the test.

5.12 Results of trialling the cloze test

As was anticipated, the trialling of cloze highlighted a number of observations. These were:

- A small number of cadets were observed to have provided more than one word per space, this was noted and it was given more attention during the major tryout. It was to be stressed that only one missing word had to be restored. Oral instructions were also to be given.
- In general, cadets performed better on passages C and D than on A and B. Relative ease of content words (lexical items) might have been the cause. As was noted earlier, passages A and B represented authentic materials, i.e. they were taken from the cadets own material. Passages C and D were simplified aviation materials, approximate with cadets own field of study but adapted for a language syllabus programme.
- The rate at which the cadets completed the four passages varied. It took them more time to complete passages A and B than C and D. The fastest cadet completed the four passages in 30 minutes, while the slowest cadet took 62 minutes.

Table (5.3) below gives the number of cadets attempting each

passage with the raw and the mean scores in percentages. The sample represent the three levels (groups) mentioned earlier.

Table 5.3: Individual and mean scores obtained by ten cadets on the four cloze tests (pilot study)

Cadet	Passage A	Passage B	Passage C	Passage D
1	9	6	12	11
2	10	9	13	12
3	6	9	11	11
4	5	8	9	12
5	7	8	9	8
6	5	8	8	9
7	4	6	7	8
8	3	5	7	6
9	5	7	5	5
10	0	0	3	6
Total				
	A	B	C	D
Mean =	5.4	6.6	8.4	8.8
Percentage =	36%	44%	56%	58.6%

- Text A was the most difficult
- All passages were retained for the major tryout.

The following Table (5.4) on the next page gives a summary of the overall scores obtained on the various subtests of the T.E.A.C. cloze, dictation, listening comprehension and grammar are given as percentages. The scores are reported in percentages. Interview assessments are reported according to the level.

Table 5.4

Overall results of the pilot study

Cadet	% Grammar	% Listening	Cloze one %	Cloze two %	Cloze three %	Cloze four %	Cloze total %	% Dict-ation 1	% Dict-ation 2	Interview
1	69.8	84.6	60	40	80	73	63.21	89	86	5
2	68.4	84.6	66.6	60	86.6	80	63.3	85.2	84.9	5
3	67	76.9	66.6	60	66.6	60	63.3	77.9	72.6	4
4	65	76.9	33.3	53.3	60	80	56.6	85.2	80.8	4
5	64.3	76.9	66.6	40	66.6	66.6	59.9	76.4	87.6	5
6	61.6	69.2	33.3	53.3	53.3	60.9	49.9	89.7	91.7	4
7	58.9	69.2	40	40	60	60	50	79.4	78.	4
8	52	53.8	53.3	60	86.6	60	64.9	69.1	71.2	3
9	50	46.1	6.6	26.6	53.3	26.6	28.2	63.2	60.	3
10	47.9	46.1	20	33.3	0	0	10.8	52.9	58.9	2
Mean	60.4%	68.4%	44%	46%	61.3%	56.6%	51%	76%	77%	

Table 5.5

T.E.A.C.'s intercorrelation (Trialling data)

	GR	LIS	CL1	CL2	CL3	CL4	DIC1	DIC2	SP
GR		0.60 P=0.061	-0.01 P=0.491	-0.37 P=0.190	0.04 P=0.462	0.20 P=0.322	-0.20 P=0.323	-0.21 P=0.309	0.70 P=0.028
LIS			-0.00 P=0.496	-0.06 P=0.438	0.25 P=0.280	0.14 P=0.371	-0.28 P-0.255	-0.09 P=0.414	0.89 P=0.002
CL1				-0.15 P=0.359	0.86 P=0.003	0.39 P=0.177	0.47 P=0.127	0.80 P=0.009	0.02 P=0.476
CL2					-0.28 P=0.258	-0.22 P=0.300	0.39 P=0.176	0.43 P=0.150	-0.23 P=0.294
CL3						0.20 P=0.316	0.36 P=0.199	0.66 P=0.041	-0.11 P=0.397
CL4							0.07 P=0.437	0.29 P=0.251	0.45 P=0.139
DIC1								0.66 P=0.041	-0.32 P=0.227
DIC2									-0.09 P=0.414
SP									

5.13 Test intercorrelations

Coeffecient values (though some are significant and others are not) as can be seen from the triangle above (5.5), were largely an experimental confirmation, and with the exception of cloze and dictation tests none of the other subtests were greatly duplicating the content of any of the other tests as represented by the low correlation indices presented in the triangle matrix above.

The speaking test on the other hand confirmed performances on grammar and listening. The speaking test correlated with grammar and listening at 0.70 and 0.89 respectively. These two correlations were significant at $p = 0.02$ and $p = 0.002$ level of significance respectively. The listening and the speaking tests are integrative in nature and the high correlation between them may well indicate that tests which are integrative in nature, tend to correlate highly with one another.

The moderate and the high correlation between the dictation test 2, cloze 1, cloze 2, cloze 3 and dictation 1 confirmed previous research findings with cloze testing that cloze tend to correlate best with proficiency measures that are integrative in nature. The correlation indices between the dictation and the cloze procedure were 0.80, 0.43, 0.66 and 0.66. The level of significance for these correlations were $p = 0.009$, 0.15, 0.04 and 0.04. However, due to the limitation of the sample (trialling the measure) these findings are still viewed as tentative and until the major try-out was conducted, these correlation indices are largely speculative.

5.14 Validity criteria

As was mentioned in the previous chapter, it had been decided to validate the T.E.A.C. against two external criteria: instructors' estimates and academic grades.

Consequently, instructors were asked to provide these two criteria. An English ability rating was designed for this purpose. They were also asked to provide their cadets end of term grades.

5.14.1 Instructors' estimates (English ability rating)

The instructors engaged in teaching the students academic or English courses were asked to supply estimates of their cadets' English proficiency on a 6-point scale. The scale was illustrated thus:

6	Excellent	Shows near native speaker ability, has no problem in expression or understanding.
5	Very good	Clearly a non-native speaker because of minor faults in English usage, but this does not handicap him in his studies; makes complete communication.
4	Good	Makes many mistakes in English but this constitutes only a minor handicap for him in his studies.
3	Satisfactory	Shows many weaknesses in English usage, but his English ability can be considered just adequate for his studies. A higher standard is desirable.
2	Weak	Shows very little ability in English both in expression and understanding and is well below a satisfactory standard.
1	Very weak	Shows no ability, no communication

Using the same guidelines illustrated above, instructors were also asked to supply estimates for three individual language skills. These skills were those actually included in the test battery. The purpose was to check on the reliability (internal validity) of the

global estimates made by instructors. The scale was illustrated thus:

Ability to understand spoken English	Excellent, Very good, Good, Satisfactory, Weak, Very weak
Ability to speak English	
Ability to understand written English	

5.14.2 Academic grades

Instructors were asked to provide the latest grade for their cadets in their own subjects of study. However, being a longitudinal study, grades in most cases were not possible until later particularly for those who were still doing English.

5.15 Conclusions on the trialling of the T.E.A.C.

The trialling of the dictation had highlighted that the length of some of the pauses between 'chunks' needed some alteration, in most cases expansion of time was necessary. This was noted and the two texts included had to be re-recorded with appropriate pauses. The re-recording took place in the Academy, and was made by a native speaker of English, who spoke with an R.P. accent. The measure was then tried on a representative sub-sample: The sample found it not a difficult task. The time needed to complete the dictation was 20 minutes, with the administrative procedures.

The interview with its four stages and scoring procedures worked satisfactorily. Interlocutors practiced the technique of conducting the interview and gained some experience in scoring this type of speaking test following guided scoring procedures which they were provided with. It was observed that once cadets were relaxed, they

responded effectively. It was found that some cadets were so keen that they made the interview as a realistic communication (see Appendix content of interview). In some cases, cadets interrupted the flow of the discourse by speaking topically. The length of the interview varied. Where it was felt cadets were keen to continue, they were not interrupted. In these cases the interview reached an optimum time of 30 minutes, however, these cases were not frequent. The minimum time was 10 minutes, and the very frequent time was 15-20 minutes. Cadets had no problems on multiple choice questions and found the grammatical structure format familiar to them. As was mentioned earlier, it is the only subtest which approximates in format only with the E.C.L. test. It was found that the most frequent time needed to complete the 73 grammatical structure test was 35 minutes but as was anticipated, it was only the slowest cadets who took more time, sometimes up to 50 minutes or even more.

The trialling of the cloze passages had led to a number of observations. It showed that some cadets provided more than one word per space. This was noted, and in the major tryout it was verbally stressed that only one word was to be supplied. It was observed that cadets whom the T.E.A.C. was tried on performed better on passages C and D than on A and B. This may well have been as a result of the nature of the deleted items or as a result of the difference between authentic (A, B) and simplified (C, D) aviation materials. However, this is a tentative conclusion on the cloze and the results are going to be fully discussed in the internal analysis of the test in the coming chapter after the major tryout. However, all four passages were retained for the major tryout, none were

either replaced or omitted.

As far as listening comprehension (the lecturette) was concerned, although it yielded very few questions (13) and might not be so reliable it had the virtue of good validity. It was therefore trialled and retained. Item analysis for the grammatical structure and listening tests were not performed until after the major tryout. In view of the small sample, neither reliability co-efficients nor item analysis were worked out, but these were worked out and included as a part of the internal analysis of the battery in the coming chapter. It was also demonstrated that rater reliability of the interviews was possible, i.e. the extent to which assessors agreed by giving the same score to an individual was great. However, reliability of the interview tests depended largely on a number of factors. Firstly, sticking to the different stages of the structured interview for all testees was no doubt contributing to reliability. Second, adherence to the scoring scale and to the description of the scale levels for each testee's achievement (performance) was another important factor in making scoring the interview a reliable measure of speaking ability.

The Test of English for Air Cadets (T.E.A.C.) as was viewed by the Academy's instructors claimed good face and content validity. The concept underlying the T.E.A.C. was approved by the instructors. In terms of content, cadets proficiency in English was measured in relation to their future needs <u>All</u> major skills needed in their future performance in English were included in the Battery. The skills tested consisted of items relevant to their course of study. So far as construct was concerned, cadets were tested in reading,

speaking and listening skills and to some extent in writing. In the speaking test they were given the chance to speak freely using whatever range of lexis and syntax they knew. In the grammatical structure test they were required to answer grammatical points which they will encounter in their textbooks. The listening tests included dictation tests and the reading tests (cloze) were actually based on literature of the type that cadets were likely to interact with.

However, in view of the limitations of this experiment, the most acceptable conclusion so far was that within the resources of the K.F.A.A., tests included in the Battery were possible, easy to administer and easy to score. All these skills are very important to success, as revealed by the results of the questionnaire.

The T.E.A.C. as now revised and trialled highlighted that the time needed to complete the original format was as follows:

Interview	-	structured interview	20 mins
Reading	-	four cloze passages, topics, authentic and simplified aviation materials	35 mins
The grammatical structure test	-	multiple choice items	35 mins
The listening test	-	dictation - two texts	15 mins
The lecturette	-		15 mins
		extra time for preparations and administration	10 mins
		Total time:	130 mins

In view of the small but representative sample of subjects none of the subtests was dropped and more reliable and valid normative data

will be gained after the major tryout. On this basis the time could be reduced to 90-100 minutes only.

5.16 Length of the T.E.A.C.

Justifications

The trial programme highlighted that the time needed to complete the T.E.A.C. in its original format may reach a maximum time of 130 minutes. Twice as much time is needed to complete it as compared with the A.L.C.P.T. (E.C.L.) which only takes up to one hour to complete.

The question which lends itself to investigation is: How could this length of time be justified? i.e. is it worth doing? and if so, why? In the discussion of this question, three issues are raised. These are:

a. The issue of reliability

b. The issue of validity

c. The issue of backwash effect

So far as the first point is concerned, there is no doubt that language testers are always concerned with the extent to which the tests they construct produce stable results. While it is maintained that the reliability of a test may be increased by:

a. Constructing effective items to eliminate or at least reduce the guessing factor, or

b. Increasing the number of distractors,

c. Reliability of a test can also be increased by lengthening a test, i.e. by increasing the number of items which are similar in nature, difficulty and homogeneity. In this respect Lado wrote:

> 'Lengthening a test with items that are homogeneous in interrelations and general level of difficulty with those of the short test, results in increased reliability.'
> (Lado 1961, p.339)

However, one also has to consider that the length should not impose fatigue on the part of the testees. A test of three to four hours may invalidate the results, therefore, a reasonable time is necessary. Luckily, or deliberately, we tried to ensure that our test should not exceed the two hour limit. It is hoped that by eliminating unsatisfactory items or choosing from duplicated tests such as the dictation or the cloze tests the time could be reduced to 90 minutes. However, two hours as compared to three or four hours needed to complete either the T.O.E.F.L. or the E.L.T.S. is very reasonable.

The second issue, the one of validity, is the whole point of the change. While this point has been and will be discussed in some detail in relation to both tests (the E.C.L. and the T.E.A.C.), we may at this point be able to conclude that the T.E.A.C. is a better device on the basis that it:

a. samples its content in relation to subsequent language use
b. samples its content in relation to needed language skills essential to later success in subsequent training
c. emphasis on both language use and usage
d. testing both receptive and productive skills

The third point, the backwash effect, is more or less related to the issue of validity as well. The argument may be presented as

follows:

In the process of test production, a language tester should not ignore the effects of his tests on the process of language teaching. It is maintained that a test is a component of the teaching/learning situation and its role should not be confined to measurements only. It is not enough that a test is viewed as a reliable and a valid measure only. It should have a positive and effective backwash effect, it should be seen to contribute to the teaching/learning context and not to distort the teaching programme as the E.C.L. does in this situation.

From this we may find ourselves more able to conclude that the T.E.A.C. in relation to language teaching, is better than the test it is going to replace.

Preliminary work for the main study was now completed. The test battery was now duplicated. The sample for whom the test will be administered was chosen on the basis of a certain criteria (see 5.8). It was also decided to validate the T.E.A.C. against two external criteria. For the purpose of instructors' ratings, an English ability rating form for use by instructors was developed (see Appendix). This rating was for the purpose of establishing the concurrent validity of the test. For predictive validity study it was decided to use success in one or more Aerosciences subjects as the criteria.

This success is represented in the cadet's end of session marks in one or more of the Aerosciences subjects. Results of the investigations will be discussed and reported in chapters six and seven. Chapter six will comprise the internal analysis of the

Battery and chapter seven will contain the validation study of the T.E.A.C.

5.17 The table below illustrates the aims of the different stages of the T.E.A.C's application. Findings are reported in the appropriate chapters.

pre-pilot first trial	pilot	final version
1. Assessment of suitable duration for taking each sub-test and the battery as a whole, and about ease of administration and scoring.	1. Investigation of the internal consistency of independent items and of sub-tests which are not based on discrete items. Analysis, i.e. includes item facility and discrimination	1. After invalid items have been elim inated a number of statistical treat-ments have been applied to data in order to verify several hypotheses, these are: a. concurrent and predictive validity of tests; b. test reliability c. factor analysis which involves 1. principal component and 2. rotation of factors d. analysis of variance e. test inter-correlations f. reliability of tests based on independent items.
2. To gain experience applying the integ-rative type of tests, the cloze, the dictation and the listening comprehen-sion test.		
3.To gain information about the suitability of the length of pauses between 'chunks' in the dictation tests.		
4.To gain experience about conducting the speaking test and assess raters' reliabilty of scoring the interview.		
5.To gain some normative data particularly on test correlations. Conclusions at this stage would only be viewed as tentative.		
* In view of the small sample neither item analysis nor reliability co-efficients were computed.		

CHAPTER SIX

TEST OF ENGLISH FOR AIR CADETS

Internal Analyses

6. Introduction

This chapter concerns itself with the regular internal analyses of the test Battery. Analyses were made for consistency of items and of subtests. It also reports on the subtests relationship, i.e. test intercorrelations in an attempt to observe performance among cadets on the different components of the test Battery. It also examines each subtest separately to observe distribution of scores and levels for the two samples. By correlational procedures this chapter attempts to report on the relationship between the T.E.A.C. and the present testing system (E.C.L.) now in use in the Academy. Through factor analysis techniques this chapter also examines the construct of language proficiency and subsequently tests the hypothesis which sees language proficiency as unitary rather than divisible. In addition, test of significance (Analysis of variance ANOVA) was employed to observe whether the two proficiency levels tested belonged to one or two populations.

6.1 Scoring and Data Handling

After the major tryout of T.E.A.C. was conducted, all test papers, with the exception of the oral interview test, were scored by the writer. All scoring was done manually. Scores, then, were taken off each test's booklets, the total score for each cadet was computed manually and each cadet's scores were put up in a profile of scores. Each profile consisted of scores, comprising 10 scores (the four cloze tests and their total, the two dictation tests, the

grammatical structure test, the listening comprehension and the total, in addition to one level for the oral interview). For the grammatical structure and the listening comprehension test (lecturette) the regular analyses were made for consistency (reliability) of tests and items.

6.2 Consistency of items in the grammar and the listening comprehension tests

Face Value and Item Discrimination

6.2.1 The Grammar Test

Extreme groups for item analysis (Anastasi 1961) were set up of 27½% each and a facility value (item facility) was computed for each item. Items were regarded as unsatisfactory if they were low on the Academy sample and had a validity index (discrimination or facility) below +.25. Of the 73 grammatical structure tests, 11 items were found to be unsatisfactory; their discriminations and facilities were: item facility [+.24, +.16, +.22, +.24, +.19, +.08, +0.14, +.08, +.24, +.20, +.19]; item discrimination [+.13, +.03, +.30, +.03, +.00, +.03, +.20, +.10, +.16, +.13].

Previous studies took +.20 as the cut-off point as a basis to reject items. If this study would take the above mentioned cut-off point, the total rejected items would only be 6. Nevertheless, the +.25 cut-off point was selected as a more stringent threshold.

6.2.2 The listening comprehension (Lecturette)

A total of four items were rejected from the listening comprehension test, not on the basis of item validity this time, but on the basis of their discrimination. These items had acceptable validity but did not have acceptable discrimination. These items

had the following discriminations: [+.0.08, +0.16, +0.12, +0.16]

TABLE 6.1

Rejected items

The Grammar Test

	Item No.	Item facility	Item discrimination
	7	+0.19	+0.13
	16	+0.20	+0.03
	26	+0.24	+0.30
	32	+0.08	+0.03
	33	+0.14	+0.00
	36	+0.08	+0.03
11 items	37	+0.19	+0.20
out of 73	41	+0.24	+0.10
	43	+0.22	+0.16
	47	+0.16	+0.13
	66	+0.24	+0.30

The Listening Comprehension (Lecturette)

		Discrimination
4 items	3	+0.08
out of 13	4	+0.16
	7	+0.12
	12	+0.16

A full item analysis for these two tests is presented in the appendix.

6.3 Test reliability

Reliability coefficients were computed for the various subtests in the battery. Measures of internal consistency for cloze and dictation were computed using the KR-21 formula. Because these tests, i.e. cloze and dictation were not based on independent items, the use of KR-21 was justified.

So far as the grammatical structure and the listening comprehension tests were concerned reliability were computed after the usual item analyses were carried out. Using equivalence

reliability Kuder-Richardson (21) reliability coefficients were computed after unsatisfactory items were rejected. The following table illustrates the reliability coefficents of the various subtests in the T.E.A.C.

TABLE 6.2

Reliability of T.E.A.C.

N = 92

Test	Kuder-Richardson Formula
Cloze total	0.84
Dictation 1	0.86
Dictation 2	0.87
Grammar	0.72 (item analysis)
Listening Comprehension	0.42 (item analysis)
T.E.A.C.	0.96

The results indicate that the T.E.A.C. and its components (sub-tests) are reliable measures of English proficiency for the cadets of the K.F.A.A. A satisfactory reliability coefficient for a pilot test would imply that it can potentially provide high reliable scores, higher than it does at this point of time.

6.4 The Reading Test (the Cloze)

89% of the total of the two samples attempted all four texts. In the time allowed, seven cadets attempted three texts only viz. 1, 2 and 3. The order in which they attempted the different texts suggested that they might not have had time to attempt text 4. Two cadets did not attempt text 2 and another two did not attempt text 1. Only one cadet did not attempt text 3.

6.4.1 Scoring procedures

The contextually appropriate word scoring procedure was used (see 4.14.6 for the characteristics given to this method). Total cloze

scores were computed. Cadets were not given any allowance if they did not attempt one text or more. The highest total score obtained was 44 out of a maximum of 60.

Table 6.3

Test statistics, means and standard deviations for level 1 and level 2 and for the two levels pooled of the cloze tests

Level	N	Mean					SD				
		CL1	CL2	CL3	CL4	CLT	CL1	CL2	CL3	CL4	CLT
1	47	6.72	7.14	8.93	8.36	31.12	2.45	1.82	1.79	1.76	5.9
2	45	3.51	4.77	5.55	4.40	18.24	2.2	2.1	2.77	2.67	7.72
1 & 2 pooled	92	5.15	5.98	7.28	6.42	24.8	2.84	2.29	2.87	2.99	9.39

The mean total score for all texts was 31.12 (51.86%) for level 1 and 18.24 (30.4%) for level 2. For the two levels pooled, the mean total score was 24.8 (41%). This relatively low percentage of the total mean was in fact affected by several factors. Firstly, some cadets obtained zeroes for each text which was not attempted and this was included in the individual's total score, consequently the "zeroes" were included in the computing of the total mean score for the whole sample. Second, the majority of the deleted items 39 (65%) out of a maximum 60 items were in fact functional items which in most cases required the restoration of the original word and this kept the individual test scores low. Third, the two levels pooled (two different proficiency levels) produced a lower mean on tbe basis that the lower proficiency level affected the higher, resulting in a relatively low mean.

6.5 Performance of the two samples on the four cloze tests

In general the two samples performed better on texts 3 and 4 than on 1 and 2. The mean indices for sample 1 were 8.9 and 8.3 for text 3 and 4 out of a maximum 15, while they were 6.7 and 7.1 for texts 1 and 2 respectively. For sample 2 the mean indices were 5.5 and 4.4 for texts 3 and 4 and 3.5 and 4.7 for texts 1 and 2 respectively. In view of the mean analyses above the most difficult text was text 1. Texts 3 and 4 were relatively easier than texts 1 and 2. However, an analysis of the nature of the deletions in each of the four passages revealed that texts 3 and 4 had more grammatical items than texts 1 and 2 had, in addition texts 3 and 4 were drawn from simplified aviation materials, while text 1 and 2 were authentic type of materials, i.e. were drawn from Aerosciences textbooks. These factors may have contributed to the ease and difficulty of each text.

6.6 Cloze texts intercorrelated

The following intercorrelations were obtained by means of product moment correlation coefficients. They report on the relationship between each of the four cloze tests. First, tables are presented, followed by discussion.

Sample 1 TABLE 6.4

Summary of correlation coefficients among the four cloze tests

	CL1	CL2	CL3	CL4
CL1	-	*0.35	*0.61	*0.22
CL2		-	*0.43	*0.31
CL3			-	*0.50
CL4				-

Pearson product moment correlation coefficient *All correlations are significant N = 47 see p.87 in the appendix for the level of significance

Sample 2 TABLE 6.5

Summary of correlation coefficients among the four cloze tests

	CL1	CL2	CL3	CL4
CL1	-	0.46	0.57	0.35
CL2		-	0.61	0.31
CL3			-	0.57
CL4				-

Product moment correlation coefficient. All correlations are significant N = 45 see p.86 in the appendix for the level of significance

Sample 1 & 2 pooled TABLE 6.6

Summary of correlation coefficients among the four cloze tests

	CL1	CL2	CL3	CL4
CL1	-	0.58	0.71	0.55
CL2		-	0.68	0.54
CL3			-	0.72
CL4				-

All correlations are significant at p = 0.000. N = 92

Tables 6.4, 6.5 and 6.6 above display the relationship among the four cloze tests. A low to moderate correlation among texts was observed indicating that each cloze test although text materials were drawn from relatively the same source, i.e. aviation sciences, might have contributed a different thing to the total cloze score and consequently a different thing to the total of the test battery. This suggested that cloze testing might in fact be text sensitive, i.e. different texts may test different aspects of language proficiency, passages which correlate moderately with one another might not test similar language skills. Within the limitation of this study, therefore, the claim which views cloze testing as an automatic and a possible measure of overall proficiency in a language might be met with the difficulty of standardizing the measure by including different passages with different types of deletions, in this case I cannot see how cloze differs from a battery which consists of several components, i.e. of several tests, testing different things reported here. It should be noted that the cloze intercorrelations confirm previous research findings (Moller, 1982) for instance, when he reported that his cloze tests correlated

moderately with one another concluding that each cloze test might in fact test somewhat different aspects of language proficiency.

6.7 Relations of cloze tests and other measures in the battery

The following tables report the relationship between cloze testing and the other components of the test battery. First tables will be presented followed by discussion. Through correlation coefficients we will observe and discuss the relationship between the cloze test and each of the other subtests introduced in the battery. These correlations represent different samples.

Tables 6.7, 6.8 and 6.9 show the changes in correlation with the other subtests in the battery.

Sample 1 TABLE 6.7

Summary of correlation coefficients between cloze scores and other parts in the T.E.A.C. with sample one

		CL1	CL2	CL3	CL4	CLT
1.	Grammar	0.42	0.39	0.48	0.37	0.56
2.	Listening	0.39	0.48	0.50	0.47	0.62
3.	Dictation 1	0.11*	0.21	0.22	0.35	0.28
4.	Dictation 2	0.22	0.12*	0.16*	0.17*	0.23
5.	Oral interview	0.50	0.56	0.51	0.44	0.67
6.	Total T.E.A.C.	0.58	0.53	0.64	0.52	0.76

see p.87 in the appendix for the level of significance

* not significant N = 47

Sample 2 TABLE 6.8

Summary of correlation coefficients between cloze scores and other parts in the T.E.A.C. with sample two

	CL1	CL2	CL3	CL4	CLT
1. Grammar	-0.054*	0.25	0.20	0.43	0.27
2. Listening	-0.027*	0.31	0.22	0.35	0.28
3. Dictation 1	-0.17*	0.08*	0.15*	0.27	0.12
4. Dictation 2	-0.21	0.11*	0.10	0.38	0.13*
5. Oral interview	0.27	0.50	0.48	0.40	.053
6. Total T.E.A.C.	0.15	0.42	0.49	0.63	0.56

see p.86 in the appendix for the level of significance

* not significant N = 45

Sample 1 & 2 pooled TABLE 6.9

Summary of correlation coefficients between cloze scores and other parts in the T.E.A.C. the two samples

	CL1	CL2	CL3	CL4	CLT
1. Grammar	0.55	0.57	0.62	0.72	0.72
2. Listening	0.58	0.62	0.64	0.71	0.75
3. Dictation 1	0.43	0.48	0.56	0.67	0.63
4. Dictation 2	0.45	0.48	0.54	0.68	0.64
5. Oral interview	0.65	0.67	0.70	0.72	0.81
6. Total T.E.A.C.	0.63	0.64	0.73	0.79	0.82

Pearson product moment correlation coefficient. All correlations are significant at the 0.000 level of significance

6.7.1 Results and discussion

When looking at the correlations with the total of the T.E.A.C., the differences among the four texts were not great. Cloze 1 and 2 approximate in the value of correlation with the total of the T.E.A.C. Cloze 3 and 4 also approximate in the value of correlation with the total of the T.E.A.C.

It should be noted that the deletion rate in all four texts was kept consistent, every 5th word deleted. However, the four texts resulted in low to moderate correlation among themselves and low to moderate correlation between cloze tests on the one hand and the other tests introduced in the Battery on the other, which indicates that different texts may well measure different factors of proficiency in English. These particular findings confirm previous research findings by Alderson (1978) and Moller (1982).

The correlation indices among the texts ranged from:

Sample 1	=	0.31 to 0.61)	all correlations
Sample 2	=	0.22 to 0.61)	were highly
Samples 1 & 2 pooled	=	0.53 to 0.72)	significant

This again indicates that the cloze tests introduced in the Battery were to some extent, testing different aspects of language proficiency.

In the present study the cloze correlated best with the grammar, the listening comprehension (lecturette) and the oral interview. This confirms previous research findings that cloze tests tend to correlate higher with integrative type of tests.

The moderate to high correlation between grammar and cloze test (values ranged between 0.55 to 0.72) confirmed Alderson's findings

(1978) when he observed that cloze testing was related more to ELBA's low order skills, grammar and vocabulary. Within the context of this study the reason of the moderate to high correlation between cloze and grammar may in fact be due to the nature of the deleted items (39 deletions out of a maximum of 60) were grammatical words, so it was not at all unexpected that the two tests would correlate higher. However, the correlational values between the grammar test and the different cloze tests vary, for sample 1, CL3 and CL1 correlated higher with the grammar than CL2 and CL4 did. With the two samples pooled, CL3 and CL4 correlated higher with the grammar test than CL1 and CL2.

The relationship between cloze tests on the one hand and the listening comprehension test, to some extent, confirmed previous research findings that cloze tends to correlate best with tests which are integrative in nature (see for example, Oller, 1974). Values ranged from as low as 0.22 to moderate 0.50 to relatively high 0.71.

However, confirming Alderson's findings (1978) cloze did not relate more to dictation than to the other subtests in the T.E.A.C. The value of correlations between cloze and dictation ranged from significant 0.21 to moderate 0.48 to relatively high 0.68 with the different samples. Significantly low to moderate to high correlations were observed between cloze and the oral interview. Values ranged from 0.27 to 81 with the different samples. This confirmed findings by Shohamy (1983).

6.8 Conclusion on the cloze

The cloze test as applied was a reliable measure. Reliability coefficient for the total score was computed resulting in a high reliability index at 0.84 (KR-21). In general, it was observed that the four texts correlated moderately with one another which may indicate that each text had the virtue of assessing somewhat different language skills. Within the limitations of this study it was observed that cloze tests vary in what they actually measure. The claim which views cloze as a possible measure of overall proficiency was not confirmed in this study. As we showed earlier, the cloze in general related more to grammar and listening comprehension. Different texts related differently to dictation. Within the limitations of this study, the integrative discrete point dichotomies was shown to be not applicable (Alderson 1978) to what cloze actually tests, because this subtest correlated similarly with both procedures.

6.9 The grammar test

This test was the first to be administered in the series to cadets. Cadets were familiar with its format "multiple choice questions". In fact, it is the only subtest which approximates with the E.C.L. in format but not in content. 92 cadets took the test. They were provided with separate answer sheets which had a maximum number of 120 questions each with four alternatives, A, B, C and D. With this test, very little time was wasted for administrative purposes.

6.9.1 Scoring procedures

One mark was awarded for each correct answer. The four alternatives were listed under each item. Cadets were instructed to darken the letter they thought was the most appropriate answer.

Sample 1 TABLE 6.10

Distributions of scores on the grammatical structure test

Scores obtained	Freq	Freq %
50-59	3	6.38%
45-49	20	42.55%
40-44	22	46.8 %
39	2	4.25%

N = 47

Mean = 44

Sample 2 TABLE 6.11

Distributions of scores on the grammatical structure test

Scores obtained	Freq	Freq %
59	2	4.44%
35-38	15	33.3 %
30-34	17	37. 7%
25-29	9	20 %
23	1	2.22%
18	1	2.22%

N = 45

Mean = 32

Sample 1 & 2 pooled TABLE 6.12

Distributions of scores on the grammatical structure test

Scores obtained	Freq	Freq %
50-59	3	3.26%
40-49	42	45.65%
30-39	36	39.13%
20-29	10	10.8 %
18	1	1.08%

N = 92

Mean = 38.4

6.10 The oral interview

The interview was not conducted until after the whole sample had taken part in all other subtests. Time consuming but worth doing, the interview was conducted last. The interview was possible with 91 cadets; only one cadet was awarded level 1. Confident cadets were fairly relaxed but with some others, interlocutors had to work hard to make those who were aware of their low proficiency in oral expression feel relaxed. In most cases it was possible.

Three native English speakers conducted and scored the interview tests. They were themselves involved in teaching English to the cadets in the Academy, so they were not complete strangers to cadets. The uniqueness of the interview was represented in the topics of discussion. Although questions were the same for everybody, apart from unpredictable questions based on cadets answers, experiences were different for individual interviewees, therefore the content of responses were different, i.e. interview

contents were not repetitive. With some cadets, the interview represented a realistic communication and yielded an interesting exchange.

The following is an example:

One of the interlocutors was asking a cadet if he had been abroad. The cadet answered 'yes', 'I have been to many countries'. He was then asked to name the countries he had been to. Making use of his knowledge of the world, he said 'I have been to your country I mean England ...'. The interlocutor then asked the cadet where about in England. He answered: 'to Exeter' then the cadet asked the interlocutor 'if he had been there'. In this case as well, the interlocutor gave the cadet the chance to ask questions. The following exchange took place:

Cadet: Before I would like to ask about your name?

Interviewer: Well, my name is Robert, my given name is Robert and my family name is Triggs.

C. And what does that mean? I mean Triggs of course. As you know it is a little bit different about the Arabic language and English language. Triggs in the Arabic language is very strange but if you have any meanings for this name.

I. Well it only means that the first three letters T-r-i come from Cornwall near Devon ... (the interviewer continued with his explanation then was asked by the cadet)

C. and, where do you come from?

I. from London.

C. from London itself ...

The phases of the interview as well as criteria of assessments were closely adhered to throughout. This no doubt contributed to the reliability of the procedures. Four cadets volunteered to have their interviews recorded. The full transcription of these interviews denoting different levels, is to be found in the appendix. Additional comments by interviewers on some cadets' performance is also to be found in the appendix.

Out of the 92 cadets who took the oral test, one cadet only was found to have no ability at all in his oral communication.

Because it is beyond the scope of this study to discuss each cadet's performance in the oral expression test, it will suffice to give a number of observations made during the interview and a discussion of the content of one of the interviews which its transcription is to be found in the appendix, since it was possible to obtain some recordings.

The assessment scale for the oral interview test and its definitions as illustrated in chapter four provided a satisfactory general description of the different performances observed.

6.10.1 Characteristics of the different levels of performance studied

Out of the seven point scale, only five levels of performance were awarded, i.e. 1 to 5. Of the sample tested, there was no existence of levels 6 and 7. This was not all unexpected since the teaching and the testing material do not directly teach and test these skills. They were largely individual cadets efforts. Perhaps level 7 is an impossibility to achieve, i.e. native speaker ability, but

level 6 if the speaking skill was directly emphasised, i.e. taught, might be relatively easy to achieve.

6.10.2 Level five

Some of the cadets who were awarded level 5 have generally been either abroad and took summer courses in English for some time in either Britain or the U.S.A. or have been to private schools in Saudi Arabia where English was taught to them right from the beginning of their elementary education. These two factors particularly the former no doubt influenced their performance positively. In general their performance was described as: effective communication. They were described as confident speakers. Their utterances were often complex and had a good range of lexical items.

6.10.3 Level four

Cadets' performance was described as adequate communication. The two main features of their oral production were occasional hesitations and impaired by minor faults of English usage. The range of their lexical choices was somewhat limited. Omissions and additions of articles were obvious characteristics of their communication. However, these faults were largely infrequent. The occasional hesitations were mainly due to searching for the appropriate structural/lexical items.

Below level 4 important changes occurred. A noticeable feature of levels 3 and 2 cadets' performance was that it was frequently slower, marked by pausing and hesitations. The range of the lexical items and the complexity of syntax were limited and reduced. Level 2 cadets in particular often produced 'yes', 'no' responses and the

pre-final phase of the interview with this level was not possible, i.e. asking questions or questions by them. Some questions were either too difficult to be understood by them or were misinterpreted.

6.10.4 Level three

The general characteristics observed were: frequent grammatical and lexical errors which sometimes obscured the meaning of the produced utterances. Simplicity of utterances in terms of syntax but when produced they were often grammatically accurate.

6.10.5 Level two

The main feature of level 2 cadets were difficulties in communication and lack of ability to pronounce basic words. Lack of confidence, hesitations, due to the limitations of the lexical items they knew.

One interviewer reported an interesting case. He attributed errors produced by one cadet as due to carelessness rather than ignorance, and he recommended that the cadet should pay more attention to his speech.

6.10.6 Approximation of interviewers and interviewees

The purpose of this section is to report on an interesting phenomenon observed during the course of transcribing a sample of the interview test. This phenomenon is known as 'approximation' in the field of language acquisition. As a strategy in interlanguage communication, approximation is defined as 'a deviant linguistic system actually employed by the learner attempting to utilize the target language' (Nemser 1974).

In our data the effect of approximation on communication is

comprehensible even though deviant. Circumlocution is one feature of learners' discourse but simplification through paraphrasing in order to facilitate input to generate interaction is a feature of interviewers' discourse.

Below four examples are extracted from the transcriptions. The first two belong to learners' discourse. They illustrate types of strategies employed by learners. The third and fourth examples belong to interviewers. They illustrate types of simplification as a kind of approximation to learners' language. Examples are first presented then discussed and the implications of approximation on teaching practices are also presented.

Example one.
Interviewer: Nine rooms in the palace! Quite a small palace?
Cadet: No, these nine rooms, its only for sleeping.
I: Nine bedrooms
C: Yah nine bedrooms

Example two:
Cadet: I have been to your country, I mean to England ... if you know Exeter ...
Interviewer: In Devon
C: Yah in Devon. It's the head of Devon I think
I: It's the capital
C: The capital
I: Yes that's right

Example three:
Interviewer: Have you ever been abroad? Have you been outside Saudi Arabia?

Example four:
Interviewer: And what did you travel by? Greyhound, bus or did you fly to different cities?

It should be noted that these are only some of the communicative strategies employed by learners during the interview tests.

In the first example, the learner either does not know the word 'bedroom' or does not remember it. Semantically the word he

produced 'room' in 'rooms are for sleeping' is more general and learners L2 always develops from general to specific.

In the second example the learner uses the word 'head' to denote 'capital'. He seems to be using a lexical substitute to fill a gap in his vocabulary.

In example three we notice that the instructor by using approximation tries to simplify his utterance through the use of a reduced lexical item. Although he did not put his hypothesis to testing to find out whether what he hypothesised is true or not he assumed that the cadet might not know the meaning of the word 'abroad' therefore he reinforced his question with another illustrative question which he thought would simplify his question.

In example four the interlocutor directs the cadet to what his message was.

6.10.6.1 <u>Implications and teaching practices</u>

One danger of using approximation in the classroom is that it may lead to misunderstanding and consequently the learning of deviant grammatical or lexical items. While I am in agreement with the view that effective language teaching implies preventing the formation of permanent intermediate systems (i.e. deviant grammatical or phonological structures for instance), nevertheless, it is evident that an approximative system on the part of the learner particularly at early stages of language learning does exist. Our role as language teachers should be directed to approximate skilfully with learners otherwise the misuse of this strategy in teaching may have dangerous consequences. Serious grammatical approximation, omission of auxiliary verbs for instance,

might result in the learning of an erroneous system. An unsuccessful paraphrase offered as explanation may lead learners to misapprehend the meaning of a lexical item, and thus later misuse it. The teacher's job is to use correct language, it may be simple but should be correct. Another danger of approximation is that the distinction between acceptable and non-acceptable usage is blurred. The learner of a target language lacking any criterion for deciding what is an acceptable or unacceptable norm for a particular usage will completely depend on the teacher, we therefore have to be very careful in employing the right strategy or a particular teaching/ learning strategy.

Samples 1 & 2 TABLE 6.13

Distributions of levels awarded for the oral interview test

*level		Freq	Freq%
5.	Full communication, minor faults	15	16.3%
4.	Adequate comunication, few weaknesses	28	30.4%
3.	Communication exists, but with frequent errors in usage	18	19.5%
2.	Limited communication	30	32.6%
1.	No communication	1	1.08%

* Levels 6 and 7 were not awarded
N = 92

6.11 Relation between the E.C.L. and the T.E.A.C.

The E.C.L. scores were obtained for the two samples who took the T.E.A.C concurrently, i.e. the T.E.A.C. was administered to the two samples one week after they had their E.C.L. test. The aim was to observe through correlation coefficients the relationship between performance on the E.C.L. and on the T.E.A.C. subtests and its total

and also in order to obtain evidence on the T.E.A.C. content validity.

Sample 1 TABLE 6.14

Summary of correlation coefficients between the different parts of the T.E.A.C. scores and the E.C.L.'s scores sample

	Grammar	Listening	Dictation 1	Dictation 2	Cloze 1	Cloze 2	Cloze 3	Cloze 4	Cloze Total	Oral Interview	TEACT
ECL	0.531	0.4051	0.36	0.426	0.17	0.37	0.46	0.39	0.45	0.31	0.57
	p=0.000	p=0.003	p=0.007	p=0.002	NS	p=006	p=0.001	p=0.001	p=0.001	p=0.01	p=0.000

Sample 2 TABLE 6.15

Summary of correlation coefficients between the different parts of the T.E.A.C. scores and the E.C.L.'s scores sample

	Grammar	Listening	Dictation 1	Dictation 2	Cloze 1	Cloze 2	Cloze 3	Cloze 4	Cloze Total	Oral Interview	TEACT
ECL	0.57	0.57	0.42	0.34	0.43	0.54	0.69	0.48	0.70	0.63	0.72
	p=0.000	p=000	p=0.001	p=0.009	p=0.001	p=0.000	p=0.000	p=0.001	p=000	p=0.000	p=0.000

As can be seen from the tables above, 6.14 and 6.15, twenty correlations between the T.E.A.C. subtests and totals on the one hand and the E.L.C. totals on the other were computed. All correlations were calculated with raw scores. Out of the twenty-two correlations only one was found to be not significant beyond the 0.0002 level of significance (see tables 6.14 and 6.15 above). As expected E.C.L. correlated best with the grammar and the listening (lecturette) tests, although correlations were low ranging from 0.40 to 0.57. These tests were similar in nature to the E.C.L. test, but content was different. The remaining T.E.A.C.'s subtests did not seem to relate to the E.C.L. more than the grammar and the listening tests did. With both samples the cloze tests correlations ranged from very low, 0.17 to 0.54 between the E.C.L. and cloze tests 1, 2 and 4. Cloze 3 on the other hand correlated with E.C.L. at 0.46 and after an examination of the deletions it was found that deletions with this test were largely syntactic words; however, the correlation remains within the low values. Moreover, E.C.L. did not seem to relate to the oral interview and the dictations as it did to the other measures introduced in the T.E.A.C. Correlations ranged from 0.34 to 0.42 with dictation, and from 0.31 to 0.63 with the oral interview. The T.E.A.C. total on the other hand correlated at 0.57 and 0.72 with the E.C.L. total with samples 2 and 1 respectively. It can therefore be concluded that language skills tested in both Batteries did not seem to relate greatly to each other. The evidence was observed through the low to moderate correlations between the two tests.

6.12 Factorial structure of the various subtests

Four types of the T.E.A.C. components can be distinguished. First, there were three sets of tests aimed at listening tasks. They included the listening comprehension test (the lecturette) multiple choice, and two dictation tests. Second there were four texts aimed at reading comprehension task (standard cloze). However, analysis of the four texts revealed that 39 deletions out of a maximum of 60 were in fact functional or grammatical, therefore the cloze tests might measure a large component of the knowledge of the English language grammar. The third type of test included aimed at measuring grammatical knowledge through discrete item tests; however, it may as well measure reading comprehension. The fourth and last type of test included, aimed at a speaking task which involves an element of listening comprehension test, in addition the method of assessment of the oral production indicates that it measures grammar, lexical items and communication (see method of assessment). The Battery therefore looks like the following:

	Grammar
Cloze tests	Reading Comprehension
	Lexical items (vocabulary)

	Grammar
Grammar Multiple Choice	
	Reading Comprehension

Dictations

Lecturette

Listening

Listening Comprehension

Grammar

Speaking

Lexical items

Communication

Generally speaking, a rough factorial structure can be identified among these tests. The cloze and the grammar on the one hand aimed at knowledge of the grammar and reading comprehension tasks, and on the other hand, the dication and the lecturette, aimed at listening comprehension tasks. The oral interview involves an element of listening comprehension but judgements, i.e. assessment on the other hand is based on global factors which involve grammar, lexis, comprehension and communication.

Two factor solutions are presented, the result of a principal components analysis is given in tables 6.16 and 6.18 and of varimax rotation method in tables 6.17 and 6.19.

6.12.1 Principal Components Analysis and Varimax rotations over the subtests of the T.E.A.C.

A principal factor solution with iterations was used to extract initial factors. All variables i.e. test components, were used for the first analysis. Principal Components Analysis extracted two factors from the data. The results of rotated and unrotated solutions are first presented in pairs of tables for each of the two sets of data, i.e. the data represents two different populations, followed by discussion.

TABLE 6.16

Unrotated factor matrix using Principal Component Analysis for the T.E.A.C. subtests. N = 47

	Factor 1	Factor 2
Cloze total	.91	
Oral interview	.83	
Listening (lecturette)	.82	
Grammar	.78	
Cloze 3	.77	
Cloze 1	.67	
Cloze 2	.66	
Cloze 4	.63	
Dictation 2	.45	.78
Dictation 1	.47	.75

TABLE 6.17

Varimax rotated factor matrix using Principal Component Analysis factor for the T.E.A.C. subtests. N = 47

Cloze total	.95	
Oral interview	.82	
Listening (lecturette)	.75	.34
Oral interview	.74	.39
Cloze 1	.73	
Cloze 2	.69	
Grammar	.68	.39
Cloze 4	.58	
Dictation 2		.90
Dictation 1		.88

TABLE 6.18

Unrotated factor matrix using Principal Factor Analysis for the T.E.A.C. subtests. N = 45

	Factor 1	Factor 2
Cloze total	.81	
Oral interview	.77	
Grammar	.73	.53
Cloze 4	.73	
Cloze 3	.72	
Listening (lecturette)	.71	.47
Cloze 2	.66	
Cloze 1	.42	
Dictation 2	.59	.69
Dictation 1	.51	.62

TABLE 6.19

Varimax rotated factor matrix using Principal Component Analysis factor for the T.E.A.C. subtests. N = 45

	Factor 1	Factor 2
Cloze total	.97	
Cloze 3	.87	
Cloze 1	.87	
Cloze 2	.76	
Cloze 4	.63	
Oral interview	.55	.53
Dictation 2		.91
Grammar		.89
Listening (lecturette)		.82
Dictation 1		.80

6.12.2 Results and discussion

Examining these pairs of tables above, it is obvious that in unrotated factors structures (Tables 6.16 and 6.18) as one would anticipate, the first factor in both cases accounted for a large proportion of the total variance. On the other hand, the additional factor did not account for a significant amount of variance as can be seen from the same tables (6.16 and 6.18) above if factors were

left unrotated. However, when the initial factors were rotated, completely different factor patterns appear. The rotated factor structures (tables 6.17 and 6.19) gave us a completely different view. Factor 1 was not so powerful as it seemed in tables 6.16 and 6.18, and factor 2 was not so weak as it appeared in tables 6.17 and 6.19.

The data presented here are consistent with almost all previous reports on the results of factor analyses. Whenever factors are not rotated in the initial factor matrix, the first factor appeared to be so strong that the researcher inferred, wrongly, the existence of one and only one general factor accounting for all the variance in the data (see Farhady 1983; Vollmer and Sang 1983). Thus it seems to us that the conclusin of the presence of one factor only (Oller 1979) and therefore the strong support of the unitary competence hypothesis as with regard to the construct of the language proficiency had been arrived at immediately after the first step of the factor analysis was completed, i.e. after the principal component process. However, the factor analysis technique is a three step process and a significant change occurs once extracted factors are rotated and the unitary competence hypothesis may no longer be supported once factors are rotated.

In view of the factor analysis results reported above, it seems to us that the strong support of a unitary competence hypothesis is not plausible and that the construct of language proficiency might in fact be more complex. The problem with principal component analytic studies is that the method of analysis is intended to simplify data, to reduce it to one solution, thus likely to result in one factor

emerging from the intercorrelations. The reason for this is that principal component analysis looks only at variance structure, not covariance structure.

6.12.3 Discussion of loadings

Below we will discuss the loading of the two factors on each of the 10 variables using the rotated factor matrix in an attempt to come up with meaningful linguistic terms for the two factors extracted. It should be noted that the three step process used were first initial unrotated factor analysis, i.e. principal component analysis, followed by varimax solutions and then the two factors extracted correlated. Although either of these steps could be used to either support or defy the hypothesis of the construct of language proficiency, the second and the third processes give more accurate information and that conclusions based on the initial analysis may not be at all complete or even acceptable.

From the results presented in tables 6.16 and 6.18, (unrotated factors), it can be seen that there is both a general language proficiency factor but not unitary and a smaller additional factor which may be labelled dictation factor. This result is anticipated in the sense that when factors are not rotated the first factor always loads on every variable, but with different values of loadings, but at the same time the additional factor also accounts for some of the variances. However, a significant change occurred once factors were rotated (tables 6.17 and 6.19). These tables represent two different samples. In table 6.17, the first factor had its highest loading on the cloze tests (the four tests together) and accounted for eight of the variables save the Dictation tests

which factor 2 had its highest loadings on them, .90 and .88, factor two also accounted for the listening comprehension and the oral interview tests. However, factor one did not align itself exactly with all eight variables but loadings values have ranged from as low as .58 (on Cloze 4) to high on the Cloze total, .95. It remains to be seen what the variables have in common. The listening comprehension, the dictation and the oral interview which have been accounted for by factor two have in common an element of comprehension, a memory span and "ability to render an immediate recall of language materials presented orally" (J.B. Carroll 1983). However, dictation is different from them in the sense that it involves some writing skills. So far as the second sample was concerned almost the same factorial patterns occurred. However, the loading values were different from sample 1, the dictation and the listening comprehension and the oral interview which have been accounted for by factor 2 had very high loadings values, .80, .82 and .91 on Dictation, Listening Comprehension and Dictation 2 respectively.

6.12.4 Factors correlated

The third step in factor analysis which is sometimes neglected in factor analysis technique is the correlation procedures among factors. This helps to observe the relation between the factors, it also indicates whether each factor stands alone or not. However, the two factors which were extracted above were correlated.

Correlation indices were:

TABLE 6.20

Level 1

	Factor 1	Factor 2
Factor 1	.91	.40
Factor 2	-.40	.91

Level 2

Factor 1	.74	.66
Factor 2	-.66	.74

6.12.5 Conclusions on factor analysis

Several conclusions can be drawn from the above analyses. First, factor analysis is a complex technique and requires several steps. These steps are complimentary to one another. Conclusion arrived at after principal component analysis can be misleading. Once factors were extracted, they should be rotated and then factors correlated. However, the conclusions arrived at within these data were:

(1) Two factors were extracted with the analyses in both the Principal Component Analysis and the varimax solutions for both samples therefore the claim which views language proficiency as unitary rather than divisible may no longer be supported

(2) Not all skills are learned to the same degree particularly with less proficient students. Some students are good at speaking but not as good at writing for instance.

(3) The interpretations of factors depend largely on the variables tested. We could therefore tentatively label the first factor as "general linguistic ability" and the second a dictation

factor with principal component.

(4) The obtained correlations of the two factors support at least the partial divisible hypothesis. While it was anticipated that each factor would correlate with itself perfectly, it did not, therefore the shape of the construct of language proficiency may well look like the following diagram:

Language proficiency

F1 F2

F1A F1B F2A F2B

(5) The extraction of factors and their composites depends largely on the components or the variables included in the battery under investigation and is affected by factors such as the characteristics of the sample and other factors. Therefore it can be expected that different test batteries applied on different samples produce different results.

6.12.6 Factor analysis continued

The factor analysis presented and discussed above included the cloze total as an additional variable. The question might therefore arise as to what extent has this variable influenced the loadings as well as the factors extracted? Since the cloze total includes cloze one, two, three and four and because these variables were also present in the analysis, we thought that this might have influenced the results obtained. However an additional factor analysis was performed this time without the inclusion of the cloze total as a distinct variable. The results of the factor analysis are presented

and discussed below. As before, results of extracted factors, rotated and unrotated are first presented followed by discussion.

TABLE 6.21

Unrotated factor matrix using Principal Factor Analysis for the T.E.A.C. sub-tests. N = 47

	Factor 1	Factor 2
Listening Comprehension	.85	
Oral interview	.85	
Grammar	.82	
Cloze 3	.72	
Cloze 2	.63	
Cloze 1	.62	
Cloze 4	.61	
Dictation 1	.51	.75
Dictation 2	.51	.75

TABLE 6.22

Varimax rotated factor matrix using Principal Component Analysis Factor for the T.E.A.C. sub-tests. N = 47

	Factor 1	Factor 2
Listening Comprehension	.83	
Oral interview	.79	.33
Cloze 3	.78	
Grammar	.77	
Cloze 1	.69	
Cloze 2	.67	
Cloze 4	.56	.90
Dictation 1		.90
Dictation 2		.90

TABLE 6.23

Unrotated factor matrix using Principal Factor Analysis for the T.E.A.C. sub-tests. N = 45

	Factor 1	Factor 2
Grammar	.85	
Listening Comprehension	.81	
Oral interview	.76	.33
Dictation 2	.74	.54
Cloze 4	.66	
Dictation 1	.64	-.49
Cloze 1		.81
Cloze 3	.56	.65
Cloze 2	.54	.57

TABLE 6.24

Varimax rotated factor matrix using Principal Component Analysis Factor for the T.E.A.C. sub-tests. N = 45

	Factor 1	Factor 2
Dictation	.92	
Grammar	.88	
Listening Comprehension	.82	
Dictation 1	.81	
Cloze 3		.86
Cloze 1		.79
Cloze 2		.77
Cloze 4	.38	.60
Oral interview	.51	.60

6.12.7 Results

The results as shown in Tables (6.21, 6.22, 6.23, 6.24) are almost consistent with the previous factor analysis discussed earlier. Principal Component analysis extracted two factors confirming the divisible hypothesis. When factors were left unrotated, factor one in both levels accounted for a large proportion of the total variance. However, once factors were rotated, a different factor pattern appears. Results of the first level were almost the same as

in the previous factor analyses reported earlier in this chapter. Dictation tests were largely accounted for by factor two.

6.13 Proficiency levels as measured by the T.E.A.C.

The two samples

Through analysis of variance technique, below we will compare performance on the T.E.A.C. by the two levels of the samples tested. First the two levels, level 1 and level 2, were compared. Second, performances on the various sub-tests are taken by the two levels are compared and differences are presented and discussed.

6.13.1 Level 1 and level 2: Second hypothesis tested

Common sense may suggest that the more proficient a student is on English language the more likely he is to perform better on an English language test aimed at measuring his ability (proficiency) than a less proficient student would do. The two levels tested in the T.E.A.C. represented two different samples. Level one comprised of cadets who had either just started doing aerosciences or were waiting to begin doing Aerosciences. They were considered more proficient on the basis of three grounds or on a combination of two of them. First, some cadets have been to English speaking countries and took summer English courses there. Second, they were considered more proficient on the basis of their score on the E.C.L., i.e. scored 75% or more. Third, they either completed an intensive English language course either in the Academy or somewhere else either in Saudi Arabia or abroad. The second level were still doing English in the Academy and still scoring below 75% on the E.C.L.

Common sense might, therefore, suggest that level 1 (the more proficient) mean performance on every test is higher than level 2

(the less proficient). Not only were the two means on every test different, but the standard deviations were different too. Level one had higher means but lower standard deviations and level two had lower means but higher standard deviations.

6.13.2 Analysis of variance (ANOVA)

A series of one way test of significance by Analysis of Variance were computed on the two levels for every test to observe whether these differences could be illustrated by chance. Table 6.25 below displays the means of the various subtests of the two levels and tables 6.26 to 6.36 presents the analysis of variance tests of significance. It should be noted that the two levels differ in the number of cases. Level 1 had 47 cases, but level 2 had 45 cases. The ANOVA technique as was used allowed for unequal cell numbers. It should be noted that t-test is also an appropriate test of significance and probably the most widely used statistical test for the comparison of two means because it can be used with small sample sizes. However, t-test is not appropriate when more than two means are to be cross-compared. For the analysis of our data either tests of significance can be used,

TABLE 6.25 Means of T.E.A.C. and its parts with the different samples

Sub-test		Level I	Level II	Total
Grammar	N	47	45	92
-	x	44.4	32.2	38.4
Listening Comprehension	N	47	45	92
	x	8.8	5.5	7.2
Dictation 1	N	47	45	92
	x	56.4	40.2	48.5
Dictation 2	N	47	45	92
	x	61.8	44.7	53.4
Cloze 1	N	47	45	92
	x	6.7	3.5	5.1
Cloze 2	N	47	45	92
	x	7.1	4.7	5.9
Cloze 3	N	47	45	92
	x	8.9	5.5	7.2
Cloze 4	N	47	45	92
	x	8.3	4.4	6.4
Cloze total	N	47	45	92
	x	31.1	18.2	24.8
Oral interview	N	47	45	92
	x	4.2	2.3	3.2
T.E.A.C. total	N	47	45	92
	x	206.6	143.3	175.7

TABLE 6.26 Grammatical structure test
One-way ANOVA source table for the grammar test on two different groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels Level 1, level 2	3410.57	1	3410.57	185.3	0.000
Residual (error)	1656.4	90	18.4		
Total	5066.9	91	55.6		

TABLE 6.27 Listening Comprehension (lecturette)
One-way ANOVA source table for the lecturette on two different groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels Level 1, level 2	249.87	1	249.8	250.2	0.000
Residual (error)	89.8	90	0.99		
Total	339.6	91	3.7		

TABLE 6.28 Dictation 1
One-way ANOVA source table for the dictation test (one) on two different groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	6067.4	1	6067.47	186.9	0.000
Residual (error)	2921.4	90	32.4		
Total	8988.9	91	98.7		

TABLE 6.29 Dictation 2
One-way ANOVA source table for the dictation test (two) on two different groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	6702.77	1	6702.77	208.8	0.000
Residual (error)	2888.0	90	32.0		
Total	9590.8	91	105.3		

TABLE 6.30 Cloze 1
One-way ANOVA source table for the cloze test (one) on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	237.27	1	237.27	42.8	0.000
Residual (error)	498.6	90	5.54		
Total	735.8	91	8.0		

TABLE 6.31 Cloze 2
One-way ANOVA source table for the cloze test (two) on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	129.27	1	129.27	33.2	0.000
Residual (error)	349.7	90	3.8		
Total	478.9	91	5.2		

TABLE 6.32 Cloze 3
One-way ANOVA source table for the cloze test (three) on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	262.77	1	262.7	48.4	0.000
Residual (error)	487.9	90	5.4		
Total	750.6	91	8.2		

TABLE 6.33 Cloze 4
One-way ANOVA source table for the cloze test (four) on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	360.87	1	360.87	70.9	0.000
Residual (error)	457.6	90	5.00		
Total	818.4	91	8.9		

TABLE 6.34 Cloze total
One-way ANOVA source table for the cloze total on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	3815.6	1	3815.6	81.3	0.000
Residual (error)	4223.5	90	46.9		
Total	8039.2	91	88.3		

TABLE 6.35 Oral interview
One-way ANOVA source table for the oral interview on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	83.1	1	83.1	237.4	0.000
Residual (error)	31.5	90	0.35		
Total	114.6	91	1.26		

TABLE 6.36 T.E.A.C. total
One-way ANOVA source table for the overall scores on two groups

Source of variation	sum of squares	DF	Mean square	F	Significance of F
Between levels	91679.9	1	91679.9	280.9	0.000
Residual (error)	29372.3	90	326.3		
Total	121052.2	91	1330.2		

6.13.3 Results

In all tests the mean differences are shown to be highly significant beyond the 0.0001 level of significance. In each test there is a level factor. This factor indicates a superiority for Level 1 over Level 2 in every test. However, this conclusion is in line with the characteristics laid down for every level. A further ANOVA was computed, this time on total score to observe whether in overall English proficiency "as measured by the T.E.A.C. (Table 6.36) these differences are intensified or suppressed" (Davies 1965).

Results of ANOVA between the two levels on total score means shown in Table 6.36 suggest that there is a significant difference in the T.E.A.C. total between Level 1 and Level 2 that can be related to their ability or level of proficiency, this indicates that the two samples belonged to two different populations.

6.14 Summary of the discussion of conclusions on the internal analysis of T.E.A.C

Several conclusions could be drawn from the internal analysis of the test Battery.

6.14.1 Internal validity and reliability of sub-tests and items

Item analysis

Discrete item tests showed satisfactory internal validity. Out of 86 independent items, comprising the grammatical structure and the listening comprehension tests, a total of 15 items were found to have unsatisfactory validity. Some items were rejected on the basis of their facility and some on their discrimination. This may have a kind of practical implication. It has been and always is the

practice of test constructors to include more of 'discrete items' or even 'integrative tests' than will eventually be needed for the final version of the test. The amount of rejected items will always depend on how effective the process of constructing the items were and the choice of distractors. However, common sense may suggest, that no matter how well one constructs his test, he will always end up with some invalid items, therefore one should right from the beginning need to include more items than he will eventually need for his test Battery.

6.14.2 <u>The Battery and various sub-tests' reliability</u>

Internal consistency of T.E.A.C. and sub-tests indicated that the Battery as a whole had a very satisfactory reliability (.96), this was anticipated on the basis of two factors;

a. the adequate internal validity of discrete item tests, and

b. the number of items and the variety of tests included in the Battery

It has already been argued (chapter five) that the longer the test is the more reliable it will show itself. The low reliability of the listening comprehension test (lecturette) may well be attributed to the total number of items in this test. 13 items were constructed for this test. 9 items only were found to be satisfactory. This may explain the low reliability index of this sub-test, at the same time confirming the view which sees the longer test as being more reliable than a shorter one. Other measures in the Battery were also reliable. Reliability coefficients ranged from .72 to .87.

6.14.3 The cloze tests

Results of correlation coefficients among the four texts ranged from low to moderate indicating that each of the four cloze texts might have tested different aspects of language proficiency. It may well be the case that the view which sees cloze as a possible measure of overall proficiency in language may well not be supported and one may need different texts with different types of deletions to claim that cloze may in fact test not overall proficiency in language but some aspects of proficiency related more to 'grammar' and 'lexical items' as two levels of the system of language since deletions are either syntactic or lexical items. In addition, it may well test reading comprehension as a language skill.

6.14.4 Correlation between listening comprehension and grammatical structure test

One striking conclusion of the inter-test correlation coefficients is the high correlation between grammar and listening comprehension. The trial programme (pilot) and the major try-out highlighted that high and significant correlation between these two tests persisted. Highly significant coefficients were observed. These were: .86 (N = 45), .90 (N = 47) and .60 (N = 8). This high correlation may well be attributed to the teaching methods of English as dictated by the A.L.C. (see chapter two for discussion).

6.14.5 Factor analysis

Different sets of data used in factor analysis technique highlighted that neither the unitary competence hypothesis nor the completely divisible hypothesis were plausible. Results of principal component analysis indicated the presence of both a

general factor but not unitary which accounted for most of the variance in the T.E.A.C. and an additional factor which aligned itself with the dictation tests. Varimax rotated solution also extracted two factors, supporting the partial divisible hypothesis. The loadings indicated that language skills may be learnt at different degrees with different levels.

6.14.6 Analysis of variance (ANOVA) and T-tests

Results of ANOVA and T-tests, tests of significance, indicated that the two levels on whom the test was piloted belonged to two different populations. Significant differences between the two levels were observed. This may well indicate that the test has discriminated between the two levels well, not only in terms of individual tests but in overall proficiency as measured by the T.E.A.C.

6.14.7 The oral interview

Reliability of the oral interview depended greatly on adherence to the stages and to the scale of assessment which raters were provided with. As was piloted, the procedures worked out satisfactorily and highlighted that assessment of oral communication can be both possible and to some extent reliable.

CHAPTER SEVEN

T.E.A.C. EXTERNAL ANALYSIS

VALIDATION STUDIES

By appealing to criteria outside the test battery that is by validity this chapter attempts to report on the validation studies made on the T.E.A.C. Firstly types of validity are mentioned, secondly the concurrent and predictive validation studies made on the T.E.A.C. are presented and discussed. In addition, face, content, and construct validity are discussed in relation to the T.E.A.C.

7.1 Types of validity

In the field of psychology testing, language is no exception, because it is concerned with assessing language as human behaviour, five types of validity are identified, these are: face, content, concurrent, construct and predictive (Cronbach 1960, Anastasi 1961). With the exception of face validity, which Anastasi called the 'superficial appearance of the measure' (Anastasi 1961), all the remaining types could be verified, either statistically and (or) theoretically.

7.1.1 Face validity cannot be established through statistical means (i.e. empirically) but can be elicited through reactions of test consumers, cadets and instructors, for instance. The test can be examined superficially and conclusions could be arrived at by asking whether the measure looks valid or not.

It is assumed that face validity is found in the T.E.A.C. The T.E.A.C. as was seen from instructors' reactions reported in chapter five do possess a tremendous face validity. The input material of

the T.E.A.C. were all related to language skills macro or micro realized in aviation contexts. However, this type of validity must not be confused with content validity. The former is not validity in the 'technical sense' (Anastasi 1961 p.104) since it does not relate to what the test is actually measuring. As with regard to face validity in the T.E.A.C. the following question may be asked.

Does the test look like an English language test for trainee pilots?

As we have seen from instructors' reactions, it is maintained that the test has tremendous face validity.

7.1.2 Another but more important type of validity is content validity. 'Content validity is an important concern for the proficiency tester' (Davies 1965, p.148).

It is regarded that the evidence essential to confirm content validity for the T.E.A.C. has already been presented in chapter four. The development of the T.E.A.C. was based on certain principles. Prior to the development of the test, we knew exactly what we wanted to measure thus content validity, in relation to skills and sampling of test material, is almost guaranteed here. The linguistic content of the individual items and of the work sample tests as was discussed earlier in chapter four provided an evidence of content validity.

It is also considered that item analysis of the grammatical and the listening comprehension (lecturette) provided a further evidence of content validity. It is also considered that a significant difference in the means of the individual subtests in the battery and in the overall language proficiency between the two levels

tested provided a further evidence of content validity. Inter-correlations among subtests (see the Appendix) of the T.E.A.C. provided a further evidence of content validity in the sense that subtests are not greatly duplicating the content of each other.

7.1.3 For the third type, construct validity, the consideration of this type of validity as with regard to language tests, raises the issue of adequate definition of "language proficiency" as well as "language", which I do not wish to address here (discussed earlier in chapters three and four) but instead it is worth illustrating briefly here the position of language testers as with regard to the importance of construct validity relating to language proficiency tests. Since the introduction of the concept of validity associated with what was termed "objective testing" the notion of construct validity has received little attention from language testers. Davies (1965: 149) noted that 'For a language test this type of validity, it is considered, is less relevant'. However in the intervening period he (1984: 52) noted that:

> 'The conclusion arrived at after this experience (experience of validation studies of language proficiency tests, since he has been involved in the construction and validation of three well known language batteries, E.P.T.B., E.L.B.A., E.L.T.S.) with proficiency tests and with validation, is that what matters above all are construct and content validity.'

However it is maintained that factor analysis indicated the existence of construct validity in the T.E.A.C. The extraction of two factors in this study provided some evidence of at least the partially divisible view of proficiency in language. At the same time, it is maintained that the various analyses of variance carried

out provided a slight evidence of construct validity of the T.E.A.C. in the sense that means (total and sub-tests) were higher for more advanced students than for less competent students.

While face and content validity of proficiency language tests made no need for the collection and empirical analysis of data, although content validity related to achievement language tests might be established through statistical means (for sampling purposes), both concurrent and predictive validity are established through empirical research.

7.1.4 Predictive and concurrent validity of language tests are checks on the value of the test (Davies 1965). Concurrent validity is established statistically by reference to some professional estimate of the ability being tested, made at the same time, hence the name concurrent, or by giving a test which its validity is known, but tests the same abilities that are being assessed in the test under validation.

Predictive validity is established by the prognostic success of a test. If a test has been designed so as to predict success in an academic or occupational setting it has to show some confirmation with what it was expected of it to do. Predictive validity study is necessarily a longitudinal one which requires a follow up, collection of data at a later stage.

7.2 Validity in the T.E.A.C.

In addition to the face, content and construct validity which have been illustrated in the previous paragraphs as with regard to the T.E.A.C., two main types of validity were thoroughly investigated. These were: concurrent and predictive.

For the concurrent validity studies, a criterion was obtained from instructors who were asked to provide estimates of their cadets' English Language Proficiency. For this purpose instructors were provided with a six-point scale (illustrated 5.18). In addition, another criterion was used, scores of the E.C.L. administered concurrently with the T.E.A.C. However it is still maintained that the E.C.L. has a negative backwash effect over the language teaching situation in the K.F.A.A. and lacks content validity as with regard to the purposes of its use at the K.F.A.A. situation but can still be used as a linguistic criterion against which the T.E.A.C. may be validated.

For predictive validity end of terms examination results in nine of the cadets own subjects were received from the head of the examination board in the Academy.

Both the concurrent and the predictive validation studies made on the T.E.A.C. will be presented and discussed below.

7.3 T.E.A.C. Validation Studies (Concurrent Validity Criteria)

All instructors whose cadets had taken part in the T.E.A.C., whether language or Aerosciences instructors were asked to provide estimates of their cadets' English proficiency at the time they took part in the T.E.A.C. The instructors were provided with a 6-point scale in which they were asked to provide both a global estimate as well as a specified estimate of their cadets' English proficiency. By specified estimates I mean the macro language skills tested in the T.E.A.C. (reading, speaking, listening) were singled out and instructors were asked to provide estimates for them separately.

The six point scale has already been illustrated (5.18).

The purpose of asking for a global estimate and an additional specified one was to check on the global estimates internal validity (i.e. reliability). However it was anticipated that instructors would give different estimates for an individual's proficiency in the three skills included, it was maintained, that specified estimates should not vary greatly from the global estimates.

For the concurrent validation studies the following returns were made and used in the subsequent concurrent validation studies.

	Global	Understanding Listening	Ability to speak	Ability to Understand Reading
	N	N	N	N
Rating 6-point scale by one instructor	15	15	15	15
Rating 6-point scale by one instructor	17	17	17	17
Rating 6-point scale by one instructor	22	22	22	22
Rating 6-point scale by one instructor	22	22	22	22
Rating 6-point scale by one instructor	8	8	8	8
N =	84	84	84	84

7.4 E.C.L. as a Criterion (Concurrent Validity)

In addition to instructors' estimates, the E.C.L. (despite reservations illustrated about it) was used to check on the T.E.A.C. concurrent validity, after all it is a language proficiency test it shares with the T.E.A.C. some skills tested. The E.C.L. was administered concurrently with the T.E.A.C. and results of both tests were used in the subsequent concurrent validity studies. However, it has already been shown (chapter two) that the E.C.L. did not relate positively to what it was expected of it to predict.

7.5 Predictive Validity Criteria

The Head of the examination board in the Academy was asked for details of academic success of these cadets who took part in the T.E.A.C. In all cases received, success represented end of term scores in nine of the Aeroscience subjects. Since the T.E.A.C. was administered in July 1985 it was not possible to obtain these results until towards the end of 1985 and others until the beginning of 1986.

The criteria supplied for predictive validation were the same for every cadet. Scores in Mathematics, Physics, Aeroengine, Aerodynamic, Flying Terminology, Avionics, Airmanship, Meteorology and Navigation were received. In addition scores which were obtained in these subjects were averaged producing an additional criterion.

The following returns were made and used in the subsequent predictive validity studies, study no. 1. It should be noted that predictive validity will be processed as results are received.

End of term scores in the subjects indicated above. Total N = 39, this total represented three classes known as 32A, 32C and 32D, which had the following numbers 14, 13 and 12 respectively.

7.6 Validation Correlation Procedure

7.6.1. Concurrent (Instructors' rating)

Through intercorrelations procedure a check on raters' internal validity of their estimates was made. The global rating was correlated with each of the skills of which criteria were provided to serve as a check on raters' reliability. However the inter-correlations were not intended to be perfect, since it was

anticipated that not all cadets have equal competence in all language skills tested, but positive and significant correlations were expected. This in a way helps reduce the unreliability factor of raters' estimates.

The tables below illustrate the intercorrelation between the global estimates and each of the specified estimates.

	Understanding Listening	Ability to speak	Understanding Reading
Global rating	.73	.50	.86

N = 45

	Understanding Listening	Ability to speak	Understanding Reading
Global rating	.96	.62	.82

All correlations are significance at the 0.000 level of significance.

The interpretation of these results is straightforward. It may be concluded that instructors rating (within this study) is to some extent reliable.

For the computation of validity correlation coefficient, 10 scores and one level for each cadet were run on the Edinburgh Regional Computer (EMAS) and the product moment correlation coefficient, between the sub-tests and the T.E.A.C. total on the one hand and the criteria on the other was produced for both concurrent and predictive validity.

Each test was treated separately in addition to the total scores on the T.E.A.C.

7.7 Discussion of Results of Validation Studies (Concurrent)

In this study as indeed in previous validation studies, for example Davies (1965), Moller (1982) and others, Cronbach's observation on the variation of validity coefficients was noted.

The variation of coefficients from group to group in this study has been noticed but this variation is not great. This might be attributed to several factors. First cadets are perfectly known to instructors, second raters involved in teaching the cadets aeroscience subjects, I came to know, have consulted their previous language instructors regarding the cadets' language proficiency. Since most cadets involved in this study have just left English classes so their English proficiency should not have changed drastically and the consultation may well have been valid particularly when it comes from professional language instructors, and thirdly the homogeneity of the cadets in terms of their language proficiency (cadets with similar language proficiency are put together) helped eliminate the heterogeneity factor which adds to the fluctuation of coefficients.

Below we will present the concurrent validity coefficients of the T.E.A.C. obtained through instructors' estimates of Cadets language proficiency followed by discussions, then results of concurrent validity, the T.E.A.C. with the E.C.L. are also presented and discussed followed by the presentation and discussion of predictive validity of the T.E.A.C. which will be discussed under predictive validity study one and two.

Table 7.1

Summary of correlation coefficients between instructors' rating and the T.E.A.C. overall scores and its parts with the different sub-sample

Concurrent Validity (Instructors' Estimates)

T.E.A.C. total with global rating

	Corr. Value	P
N = 15	0.82	0.000
N = 17	0.57	0.009
N = 22	0.73	0.000
N = 22	0.70	0.000
N = 8	0.73	0.02

T.E.A.C. Subtests with global rating

Global rating N = 15

Grammar	Listening	Dictation 1	*Dictation 2
.75	.81	.66	.37

Cloze 1	*Cloze 2	Cloze 3	*Cloze 4	Cloze Total
.51	-.08	.58	.30	.52

* Oral interview

.40

With the exception of the asterisked values all remaining correlations are significant.

Global rating N = 17

Grammar	Listening	Dictation 1	*Dictation 2	
.67	.53	*.14	*.28	
Cloze 1	*Cloze 2	Cloze 3	Cloze 4	Cloze Total
*.32	*.31	.49	.49	.48

Oral interview

.70

With the exception of the asterisked values all remaining correlations are significant. See Appendix.

Global rating N = 22

Grammar	Listening	Dictation 1	*Dictation 2	
.76	.73	.45	*.17	
Cloze 1	*Cloze 2	Cloze 3	Cloze 4	Cloze Total
.47	*-0.004	.56	*.31	*.48

Oral interview

.36

With the exception of the asterisked values all remaining correlations are significant. See Appendix.

Global rating N = 22

Grammar	Listening	Dictation 1	*Dictation 2	
.74	.54	.53	.48	
Cloze 1	*Cloze 2	Cloze 3	Cloze 4	Cloze Total
*.33	.49	.68	*.31	.65

Oral interview

.69

With the exception of the asterisked values all remaining values are significant.

Global Rating with Sub-tests

N = 8

Grammar	Listening	Dictation 1	Dictation 2	
.82	.72	*.44	*.53	
Cloze 1	*Cloze 2	Cloze 3	Cloze 4	Cloze Total
.23	*.42	.58	*.29	*.44

Oral interview

*.57

Specified rating with relevant

Sub-tests N = 15

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading material (Reading Comprehension)	.59	*.27	*-0.20	.46	*.27

	Cloze Total
	.33

	Oral Interview
Ability to speak	.55

	Listening	Dictation 1	Dictation 2
Understanding listening	.41	.47	.50

Specified rating with relevant

Sub-tests N = 17

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading material (Reading Comprehension)	.64	.48	.42	.65	.58

	Cloze Total
	.63

	Oral Interview
Ability to speak	.56

	Listening	Dictation 1	Dictation 2
Understanding listening	.45	.18	.19

Specified rating with relevant

Sub-tests N = 22

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading	.47	*.57	*-0.31	.72	*.13

Cloze Total

*.21

Oral Interview

Ability to speak .57

	Listening	Dictation 1	Dictation 2
Understanding listening	.53	*.30	*.21

Specified rating with relevant

Sub-tests N = 15

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading material (Reading Comprehension)	.59	*.27	*-0.20	.46	*.27

Cloze Total

.33

Oral Interview

Ability to speak .55

	Listening	Dictation 1	Dictation 2
Understanding listening	.41	.47	.50

Specified rating with relevant

Sub-tests N = 22

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading material (Reading Comprehension)	.69	.42	*.30	.53	.44

Cloze Total

.60

Oral Interview

Ability to speak	.84

	Listening	Dictation 1	Dictation 2
Understanding listening	.59	.52	.41

Specified rating with relevant

skills in the T.E.A.C. N = 8

	Grammar	Cloze 1	Cloze 2	Cloze 3	Cloze 4
Understanding reading (Reading Comprehension)	.61	.73	.67	.92	.48

Cloze Total

.79

Oral Interview

Ability to speak	.71

	Listening	Dictation 1	Dictation 2
Understanding listening	.74	.92	*.48

With the exception of asterisked values all remaining correlations are significant. See Appendix

7.8 Discussions of results of concurrent validation studies

(Raters' estimates)

When looking at the global rating with the T.E.A.C. total the

variation (fluctuation) is not great from one group to another. Of the first group, three had almost identical correlations ranging from .70 to .73, in fact two groups with different (NS) N = 22, N = 8 had exactly the same correlation values .73. The remaining two groups had .82 and .57 correlation values. In contrast with previous validation studies using rating or estimates as a criterion, the fluctuation here is not great. There are some explanations for great fluctuation of previous studies. In the case of the T.O.E.F.L. and the E.P.T.B., for instance, reasons behind great fluctuations have been attributed to sampling fluctuations, differences in course content, differences in reliability of grading and so on. The less variation in correlation coefficients in this concurrent validity study might be attributed to the following factors: homogeneity of groups, and of raters. However when looking at the global ratings with individual tests in the T.E.A.C. the situation is somewhat different.

7.9 Individual Tests Correlations (Concurrent Study Ratings Criteria)

7.9.1 Grammatical Structure

In the obtained correlation coefficients, this test does not fluctuate greatly. Obtained values ranged from .67 to .82. The .82 value is the odd one out and could be attributed to the total (N = 8) of this group. It was easier for the instructor who rated them to give a more accurate judgement given the low total N. The remaining correlation ranged from +.53 to +.81 with no great variation.

7.9.2 The Listening Test

With this test the values of the obtained correlation ranged from +.53 to +.81 a greater fluctuation than with the grammar test. However the variation with this test is not greater than +.30.

7.9.3 The Dictation Test 1

With this test there is a great variation indeed. Obtained values ranged from +.14 to +.66.

7.9.4 Dictation 2

The concurrent peak for this test with raters' estimates is .53. The range of correlation coefficient is between +.19 and +.53.

7.9.5 The Cloze Test

Result of the correlation coefficients of the cloze tests with the global rating will be presented together. So far as cloze 1 is concerned the range of correlation coefficients fall between +.23 and +.51, however out of the five sets of correlation coefficients three values are not significant, the other statistically significant correlation coefficients are .51 and .47. Great variation with cloze 2 on the other hand was observed. Results of correlations with this test fluctuates between a bizzare coefficient of −0.004 to +.49.

As for cloze 3 all correlation coefficients were significant. Values with this test ranged between +.49 and +.68. Cloze 4 correlation coefficients ranged between +.29 and +.49 of which only one value was significant. The cloze total fitted much better with the criteria than individual cloze tests. Correlation coefficients fall between +.44 and .69 of which two values were not significant.

7.9.6 The Oral Interview

The concurrent peak with this sub-test is .70. Coefficients fluctuate about .70 among various groups.

7.10 Discussion of the concurrent validation results

In general there is no variation greater than .4 for any individual test (except the oral interview) and the total significant coefficients obtained were higher than the non-significant ones. A totl of 29 significant coefficients in contrast with 21 non-significant coefficients was obtained.

So far as the T.E.A.C. total with global rating is concerned, all values obtained were positive and highly significant. We can draw two conclusions on the value of correlations obtained so far. Firstly that fluctuations in the coefficients were observed and this may be among other factors due to raters unreliable estimates and to the different levels of cadets' proficiency. Secondly that individual tests are not expected to show good concurrent validity with instructors' rating, not even required that individual tests should possess good concurrent validity (except that if the criterion is a similar language test).

We will now present and discuss the concurrent validity of the T.E.A.C. with the E.C.L.

7.11 Concurrent validity the T.E.A.C. with the E.C.L.

In spite of the reservations on the E.C.L. as expressed in the introductory chapter in this thesis, it was used to check on the present test's (T.E.A.C.) concurrent validity. After all, the E.C.L. is a test aimed at language proficiency. It shares with the T.E.A.C. some language skills, reading and listening. However

coefficients were not expected to be high but moderate particularly with the grammatical structure test included in the T.E.A.C., which is the only test that approximates with the E.C.L. in format but differs in content.

First results of correlation coefficients are presented then discussed.

	Grammar (T.E.A.C.)	Listening Comprehension (T.E.A.C.)	N
E.C.L.	.53	.40	45
E.C.L.	.57	.57	47

All correlations were significant beyond the 0.001 level of significance.

The E.C.L. total scores were correlated with each of the T.E.A.C.'s subtests. Although the E.C.L. had two sections (listening and reading), it was not possible to obtain individual parts' scores separately.

What is actually employed in the correlations is the total score which may run the risk of being biased, however the E.C.L. total will be correlated with the T.E.A.C. total and it is hoped that in a way the bias in correlations is eliminated. But even with this the total items are different in both tests.

From the above correlations, as was anticipated, positive but moderate correlations were obtained between the E.C.L. on the one hand and the grammar and the listening tests of the T.E.A.C. on the other.

Although the dictation may well be considered as a listening task, among others of course, the E.C.L. did not relate to it as it did to the listening comprehension test. Correlation coefficients

between the E.C.L. and the Dictation ranged between .34 and .42.

With the cloze tests the range of correlations with the E.C.L. fall between .17 and .69. This was not surprising since the systematic deletions in the cloze resulted in 39 deletions out of a maximum 60 being syntactic words.

The T.E.A.C. total with the E.C.L. total correlated at .57 and .72 with the two levels of proficiency.

We may then conclude that the T.E.A.C. is as valid concurrently as the well established test the E.C.L. but the T.E.A.C. remains superior in the sense that it has a tremendous face and at least a good content validity for the Academy's purpose.

7.12 Conclusion

Within the limitation of this study, it is therefore concluded that the T.E.A.C. possesses satisfactory concurrent validity as was validated against instructors' estimates of cadets' language proficiency and the E.C.L. which also aimed at measuring cadets' language proficiency.

7.13 Predictive validity study of the T.E.A.C. study one

By the end of February 1986, the writer received end of term exams results, (end of intermediate phase of aerosciences training). The results represented scores for 39 cadets out of the 92 who took part in the T.E.A.C. in July 1985. These scores were for the following subjects:

Mathematics, physics, dynamics, aeroengine, avionics, airmanship, navigation, meteorology and flying terminology.

The 39 cadets have taken their speciality examinations in different periods. The following table illustrates:

14 cadets completed their exams on the 20th of October 1985

13 cadets completed their exams on the 14th of January 1986

12 cadets completed their exams on the 16th of January 1986

All these results were used in establishing the predictive validity of the T.E.A.C. representing 9 criteria. An additional criterion was used and this represented the average of each cadet's scores in the nine subjects.

7.13.1 The criteria and factors affecting success in English medium classes

The main concern of a non-native student studying in the medium of the English language is to follow his speciality and complete it successfully. However a number of factors may affect success in any one speciality, among them no doubt proficiency in the language concerned. Other factors affecting success as well are the following:

a. Knowledge of the subject matter

b. Adequate instructions

c. Intelligence

d. Motivation

Adequate proficiency in a language used as a medium of instruction is fundamental to later success in the part of the candidate. Language is necessarily a condition to academic success. After all candidates are instructed in language, they read and write language, they are tested in language. However the prime consideration of a language proficiency test is not to determine who and who is not going to pass that speciality taught through that language which is being assessed but is used as an instrument to assess the adequacy

of ones language proficiency to follow a speciality conducted in the medium of that language being assessed.

7.13.2 Predictive validity study of the T.E.A.C

Predictive criteria/problems

As with concurrent validity criteria, problems with predictive validity criteria do exist. For predictive validity a lapse of time is necessary between giving the measure and the criteria sought. However if the lapse of time is quite long, a candidate's proficiency may be affected either positively or negatively. Another problem is the use of non-linguistic criteria which normally consists of language and something else.

With the T.E.A.C.'s predictive criteria the first problem may not exist at least for the 39 cadets, whose results have been received and used for this part of the predictive validity study.

7.13.3 Results of Correlations (predictive study)

The subtests' scores of the T.E.A.C. as well as the T.E.A.C. total scores were correlated with scores on each of the nine subjects which have been mentioned earlier in this chapter, an additional criteria was used and it represented the average of all the nine subjects' scores.

The Pearson product moment correlation coefficients was computed resulting in a total of (109) values of correlation coefficient for every sub group (see Appendix).

The tables below present the correlation coefficients of the T.E.A.C. total and the T.E.A.C. subtests with each of the 10 criteria.

Table 7.2

Summary of correlation coefficients between the T.E.A.C. overall scores and its parts with Aerosciences scores

		Study 'I'	
N	Range of Corr. Values	No. of Sets	Remarks
14	+.48 to +.70	48	[1] Highly significant
	+.37 to +.40	13	[2] Almost significant
	+.12 to +.36	48	[3] Not significant
13	+.46 to +.78	44	Highly significant
	+.39 to +.45	9	Almost significant
	-0.0009	56	Not significant
12	+.48 to .73	14	Highly significant
	+.41 to .47	10	Almost significant
	-0.06 to +.40	85	Not significant
39	+.26 to +.55	59	Highly significant
	+.20 to +.25	13	Almost significant
	Below +.20	27	Not significant

Table 7.3

Summary of correlation coefficients between the T.E.A.C. overall scores and Aerosciences scores

	N	N	N	N
	14	13	12	39
T.E.A.C. with G.P.A.	.56	.53	*.30	
Math	.61	*.41	*.24	.39
Physics	.38	.66	*.05	.33
Dynamics	.63	.55	*.12	.33
Aeroengine	.47	.35	*.08	.29
Avionics	*.26	.69	*.22	*.11
Airmanship	*.21	*.41	*.27	.32
Navigation	*.22	*.34	.48	.33
Meteorology	.49	.45	*.36	.43
Flying Terminology	.47	.55	.56	.51

With the exception of the asterisked values all remaining values are highly significant.

7.13.4 Discussion/Predictive Validity

An examination of the two tables presented above reveals to us a number of observations. Firstly the T.E.A.C. total with Flying Terminology as a criterion had all its correlation values significant with all three subgroups and with them all pooled

significant as well. However this is not surprising, flying terminology is the only criterion which involves heavy elements of linguistic factors, mainly E.S.P. lexis plus of course something else, and is taught by instructors as a language subject rather than an aeroscience subject. In addition this criterion is the only criterion nearer to the administration of the T.E.A.C. The first subject which cadets were examined in is flying terminology, after passing the E.C.L. These correlations ranged from .47 to .56 and were highly significant.

The second observation is that with the third group N = 12 all coefficients save two were low and not significant. However, this group was the last to be examined in aeroscience subjects, the low and insignificant correlations obtained may be attributed to the length of the lapse of time, their language proficiency may have changed in the intervening time. The third observation is that the predictive peaks of some of the T.E.A.C. subtests reached a very satisfactory level up to +.78 so it may be worth at this point examining closely each subtest in the T.E.A.C. separately in relation to each of the ten criteria, since it may be the case that the work sample tests included in the T.E.A.C. had its peak coefficients with criteria which had its input material from.

7.14 Individual test validity coefficients (predictive validity)

Every sub-test is discussed with all subgroups. Total N = 39.

7.14.1 Grammatical Structure Test

With the first group N = 14 a total of six coefficients out of 10 (including GPA as a criterion) were significant. With the second group N = 13 again six coefficients were also significant. With the

third group N = 12 only two coefficients were significant. Significant correlations ranged between +.37 to +.68. The highest correlation was with flying terminology at +.68. As we indicated earlier the flying terminology subject is one of language rather than an Aeroscience.

7.14.2 The Listening Comprehension (Lecturette)

This test showed significant and relatively high correlations with Navigation, Flying Terminology and Avionics. Coefficients ranged between +.60 and +.63 with these tests and correlated with meteorology at +.44 and .47. However, it is not at all surprising that this subtest correlated highly with most of the Aeroscience subjects save Mathematic and Physics which do not require an extensive use of language.

7.14.3. The Dictation Tests 1 and 2

These are again work sample tests. Coefficients with these tests ranged between +.66 and +.72. However it correlated significantly with a number of the criteria, with Dynamics, Math, Meteorology and Flying Terminology.

7.14.4 The Cloze Tests

The cloze tests seemed to be the best predictors of the T.E.A.C. This was not at all unexpected. Aerosciences involved a lot of reading material. The cloze was introduced in the test battery, as a measure of reading comprehension which had its input material from aviation related matter. With the first group N = 14 the cloze total, which involved cloze 1, 2, 3 and 4, correlated significantly with 7 of the criteria. Coefficients ranged between +.53 and +.70. With the second group N = 13 significant coefficients ranged between

+.51 and +.75. With the third group N = 12 coefficients ranged between +.53 to +.69. However individual cloze tests had higher predictive peaks with different criteria (see Appendix).

7.14.5 The T.E.A.C. Total

On the whole the T.E.A.C. total correlated significantly and positively with almost every criteria used to establish the predictive validity of the T.E.A.C. Significant coefficients ranged between +.47 and +.69, not mentioning correlations which are in the +.40's but are almost statistically significant.

7.14.6 Conclusion on predictive validity

The main conclusion to be drawn at this point is that predictive validity for the T.E.A.C. is claimed within the limitation of this study. Not only the T.E.A.C. as a whole but some subtests in the T.E.A.C. have also acted as good predictors. However what one looks for is a positive correlate of the total of the Battery with the criteria. If on the other hand subtests as well correlated positively with the criteria it is an additional virtue.

7.14.7 Relatively High Correlation: T.E.A.C. and Criteria Explained (Predictive)

Satisfaction with language proficiency tests correlating with predictive criteria has been expressed if correlations reached a value of .3. Typical predictive correlations are about .3, (Davies 1984), or at least this is what previous correlation coefficient studies have indicated. The question which lends itself to investigation, here, is how may the relatively high coefficients obtained between the T.E.A.C. and criteria be explained?

English medium courses' instructors have been expressing concern

with regard to non-native speakers' language proficiency. Moreover, since the early sixties language proficiency test designers have been interested in correlations of their language tests with subsequent academic outcomes. These points take us back to the question: how would one anticipate a language proficiency measure to correlate with other measures aimed at assessing knowledge of a subject matter rather than language? Aerosciences for instance.

As we noted earlier a number of variables influence academic outcome, among them no doubt language proficiency (this is of course with reference to non-native speakers). It could be safe to conclude that a more proficient candidate in a language is likely to do better in his exams than a less proficient candidate studying something other than language. However, a number of explanations are behind the relatively high coefficients between the T.E.A.C. and what it was decided to predict or better correlate positively with.

Firstly, in contrast with well established language proficiency tests such as the E.P.T.B. and the T.O.E.F.L., the T.E.A.C. was constructed for a more homogeneous group (special group) but above all was validated against "unified criteria". This special group all having:

a. a shared L_1 (all speakers of Arabic)

b. the majority of them underwent the same L_2 language programme, prior to joining the Academy.

Secondly, but perhaps, more importantly is that the T.E.A.C. samples its content in language contexts related to all testees experience therefore satisfies all, however it should be noted that the T.E.A.C. does not account for individual differences, but it deals

with a desired proficiency.

The following matrix illustrates differences between the T.E.A.C. on the one hand and the E.P.T.B., the E.L.B.A. and the T.O.E.F.L. on the other which may explain the T.E.A.C. coefficients obtained.

		T.E.A.C.	E.P.T.B., T.O.E.F.L. etc.
Target population) Testees		a. Shared L_1 b. Same target language background c. Identical speciality	a. Different L_1s b. Different language background c. Different specialities
Test content and proficiency types		a. Receptive and productive skills tested b. Discrete, integrative items c. Topics related to subsequent training d. Linguistic, overall and communicative e. Structural/ psychometric and psycho/socio linguistic	a. Largely receptive skills b. Discrete and integrative items c. Topics of general interests (gross differences) d. Largely linguistically based e. Largely structural and psychometric
		The same for all	Different for all
validity criteria	concurrent	a. Instructors' estimates b. Language test	a. Instructors' estimates b. Rank order
	Predictive	a. end of terms exams results.	a. Scores and pass/ fail points.
		The same for all	Different

7.15 Cut Off Score/Decision Making

Correlation coefficients (predictive or concurrent) are not useful in deciding what a cut-off score for a particular situation might

be. However, a decision has to be reached as with regard to the use of the T.E.A.C in the K.F.A.A. for selection purposes. For this purpose expectancy tables were set up. They illustrate the relationship of the T.E.A.C. total score to concurrent validity criteria and predictive validity criteria.

7.15.1 Concurrent Validity

In the case of the concurrent validity scores where the criterion was based on a 6-point scale estimate, the 'pass' level for the criterion was decided to be 3, scores 1 and 2 were considered to be 'Fails' while scores 3, 4, 5 and 6 were regarded as 'Successes'. It should be noted that the absence of level six was because it was not awarded to any one cadet. The description of level '2' on the 6-point concurrent scale stated that communication was limited to answering 'Wh' questions with serious grammatical and language usage errors. Where the criterion was another language test (E.C.L.), the 'pass' mark for the criterion was decided on the basis of the pass mark decided by the Academy as a pass score 75. This may seem a harsh decision for a number of reasons. First in order for one cadet to get 75 E.C.L. mark, he has to score something like 78 out of a maximum 100. This is a high and an arbitrary cut off score indeed and it may not be efficient on the basis of cost benefit effect. Secondly, as we argued earlier the E.C.L. is almost exclusively based on lexical items and some cadets may not have the potential for learning vocabulary items (i.e. ability to memorize items).

A decision had to be made with regard to the cut off score for the speaking test separately. For this purpose, the scattergram is

based on the concurrent rather than predictive validity. Since speaking may not be important to success in aeroscience subjects but is considered more important at a later stage, while flying training sessions are in progress. It should be noted that once cadets start aerosciences, no more English classes are given to them.

7.15.2 Predictive Validity

Success on the predictive validity was again determined by the Academy's cut off score on aerosciences. 60% was considered a pass, below that was considered a fail.

The scattergrams to be discussed here are those using T.E.A.C. total score with each of the criteria.

7.15.3 Different Cut Off Scores

Having demonstrated that the T.E.A.C. is a reliable and a valid measure, an attempt was made to compare different cut off scores for both the concurrent and predictive validity. First predictive validity cut off scores will be presented followed by discussion then concurrent validity follows.

7.15.4 Predictive Validity

For the predictive validity criteria the following table sets out the percentage success at scores (on the T.E.A.C. total), 177 (60%) and 194 (65.7%) to test the hypothesis that different subjects (Aerosciences subjects) needed different cut off scores.

Table 7.4

Percentage success in expectancy tables

Predictive validity criteria

Total T.E.A.C. Scores

Subjects	60% 177	65% 194
Mathematics	64%	61%
Physics	64%	61%
Dynamics	66%	64%
Aeroengine	66%	64%
Avionics	69%	64%
Airmanship	69%	66%
Navigation	69%	66%
Meteorology	69%	66%
Flying Term	66%	56%

As can be seen from the table above, the first observation is that there is no significant difference among the different subjects in terms of the percentage of success which may indicate that different aeroscience subjects may not need different cut offs. However, as can be seen from the table above there is little to choose between cut offs 177 (60%) and 194 (65.7%). The two cut offs provided almost the same percentage of success. It should be noted that the choice of 60% as a cut off point was based on the Academy's cut off score on the Aerosciences subjects.

From the point of view of predictive validity, then, either 60% or 65% would achieve similar optimum success with all aeroscience subjects. In view of the analysis above two points are raised, the first is that a lower cut off 40% or 50% may be inefficient because some cadets whose English is just adequate "bare pass" in the T.E.A.C. are admitted, given that their knowledge of the subject

matter is adequate they may encounter some linguistic difficulty and they may be at risk. On the other hand, since there is not any significant difference between the 60% and the 65% as cut off, and 60% allowed for nearly (70%) success, a higher cut off would be undesirable on the basis of their score on the linguistic criteria creating a situation of inefficiency on the basis of a cost-benefit effect, after all, linguistic ability is only one factor influencing later success.

7.16 Predictive validity study two

By the end of June 1986, the writer received end of term exams results (end of intermediate phase of aeroscience training for some cadets and results of intermediate progress tests for others.

The results represented scores for 27 cadets out of the 92 who took part in the T.E.A.C. in July 1985. The total number of cadets' results received so far is for 66 cadets. The results represented scores for the following subjects: Mathematics, Physics, Dynamics, Aeroengine, Avionics, Airmanship, Navigation, Meteorology and Flying Terminology.

The 27 cadets have completed their speciality examinations in different periods. The following table illustrates:

12 cadets completed their end of term exams on the 19th March 1986

10 cadets completed their end of term exams on the 19th April 1986

5 cadets completed their progress tests on the 16th June 1986

Each subject's scores was used as a criterion against which the T.E.A.C.'s results were validated. It should be noted that problems with the use of predictive validity criterion if it is non-linguistic, the factors contributing to success in a foreign

language medium of instruction, procedures for computing predictive validity coefficients and fluctuations of coefficients have already been discussed earlier in this chapter and there is no need for repeating them here. What, therefore, remains to be seen is the extent to which results (validation results) of this subsample reinforce the claim which has already been made for the T.E.A.C. in terms of predictive validity.

7.16.1 Procedures, results and discussion

The Pearson product moment correlation coefficient was computed between the T.E.A.C. total and each of the criterion. The table below displays results of correlation coefficients for every subgroup:

Table 7.5

Summary of correlation coefficients between the T.E.A.C. overall scores and aerosciences scores study two

	N 12	N 10	N 5	N 22
T.E.A.C. total				
with Mathematics	.56	*.42	*.59	.50
Physics	*.38	*.31	*.64	.36
Dynamics	*.44	.65	-	.51
Aeroengine	*.48	*.45	-	.47
Avionics	.60	.66	-	.67
Airmanship	.74	*.29	-	.42
Navigation	.59	*-.01	*.76	*.24
Meteorology	.73	.55	*.29	.61
Flying terminology	.74	.70	*.43	.72
Average	.61	.44	*.63	.52

\- no criterion was obtained
* with the exception of asterisked coefficients all remaining coefficients are significant

A close examination of the coefficients obtained reveals a number of observations. As before result of coefficients fluctuate from one criterion to another. Coefficients fluctuate about +.50. The

coefficients obtained ranged between .24 and .74 with different criterion. The second observation is that the T.E.A.C. correlated significantly best with Flying Terminology (Ft). The same explanations given previously for this significance may well be given here. Flying Terminology is the first subject close to taking the proficiency test (T.E.A.C.). After learning English, cadets are introduced to aeroscience by teaching them Ft which mainly consists of heavy elements of language. However the T.E.A.C. total has correlated significantly with Airmanship and Meteorology. Coefficients were .74 and .73 for these two subjects respectively. For the last group N = 5, despite high coefficients all were not significant and this may well be attributed to the small total of cases involved.

7.16.2 Conclusion

Reinforcing previous predictive coefficients, it may be concluded that there is a satisfactory positive relationship between the T.E.A.C. and what it was expected of it to predict. It should be noted that results for the remaining 26 cadets out of the 92 cadets who took part in the T.E.A.C. will not be possible to obtain in the near future because they will not start their training in aerosciences until some time later (six months from now) because some cadets are still doing English and it would not be possible to include their results in this study.

7.17 The validity studies conclusions

The results of both the concurrent and predictive validity studies indicated that the T.E.A.C. may be considered, for the levels tested 1 and 2 who represent the Academy's cadets, as a valid language

proficiency measure. The evidence provided for this conclusion is that significant coefficient ranged between +.57 and +.82 for the concurrent validity (instructors' estimates) and .57 and .72 with another established language proficiency test that is the E.C.L. (despite reservations outlined about it earlier).

So far as the predictive validity is concerned the fact that most significant coefficients were above .47 reaching a maximum of .78 with some of the T.E.A.C. sub-tests and up to 69% of the T.E.A.C. total with the criteria are all indicative to the test predictive power. Both concurrent and predictive validity were claimed for the T.E.A.C.

In addition, evidence has been presented for the existence of face and content validity in the Battery. Moreover results of factor analysis revealed that there exists a construct validity of some kind in the T.E.A.C.

The T.E.A.C., it was therefore regarded, is a valid measure of the English proficiency of the cadets of the K.F.A.A. within the sample tested and of those cadets in the two levels tested. However, the test may be used by other comparable flying trainee cadets anywhere in the world provided that they study through English, since basically Aerosciences are the same whether they are taught in Saudi Arabia or anywhere else.

7.18 Major Hypothesis Tested

It was therefore regarded that the major hypothesis that there was no positive relation between English proficiency and either instructors' estimates or academic success in various subjects, could not be supported, and the null hypothesis was therefore, on the

basis of the evidence presented earlier, rejected.

CHAPTER EIGHT

SUMMARY, CONCLUSIONS AND IMPLICATIONS

8.1 Content

This project was set up in response to the need for a new English language proficiency test and subsequently a new English language syllabus to replace the existing language programme now in use by the King Faisal Air Academy, in Saudi Arabia. In addition to the negative back-wash effect of the E.C.L. discussed in the introductory chapter, it should be noted that the A.L.C. and the E.C.L. have neither been specifically designed for cadets who will undergo aviation training in Saudi Arabia in general, nor were they intended for the Academy's cadets in particular, but were both designed for military personnel who will undergo training in the U.S.A.

After preliminary analysis and trials, a Battery of five tests (The Test of English for Air Cadets: T.E.A.C.) was designed. The sub-tests included in the Battery were based partly on discrete item linguistic categories and partly on work sample material of the kind the cadets are going to deal with in subsequent phases of training. The research on T.E.A.C. started with several assumptions and hypotheses, as indeed with previous work on language proficiency testing. Making use of advances in language testing theories the Battery comprised the following sub-tests:

Skill		Test Content	No. of items
Listening	(a)	Listening Comprehension lecture on meteorology	13
	(b)	Dictation:	68
		Dictation:	73
Reading	(a)	Grammatical Structure	73
	(b)	Cloze:	
		Text 1: Authentic Text 2: and simplified Text 3: aviation Text 4: materials	
Speaking		Oral interview	7 levels
		Total taken by each cadet + the oral interview	287

8.2 New Syllabus specifications

In addition to the design of the test Battery a set of specifications for a new syllabus was also set up. The shortcomings of the A.L.C. were presented and discussed and an alternative framework was presented and discussed. It is also hoped that the section on the project's implications would also put the proposed framework for the syllabus in perspective.

8.3 Sample

The full Battery was tried out on 103 cadets. This total represented two different populations, as was indeed confirmed by the Analysis of Variance Technique (test of significance). The two samples represented two distinct levels of proficiency. However, only 92 cadets' results have been included in the analysis. The remaining 11 cadets have either taken part or parts of the Battery, due to sickness or being away for pre-flying medical checks or have

been suspended from training in due course, therefore, it was difficult to make a follow-up study on them. However, it should be noted that these cadets were the total numbers available in the Academy who conformed with the characteristics laid down earlier but briefly repeated here for easy reference.

8.4 Characteristics of the sample

(a) Cadets who have had their English proficiency assessed by the E.C.L. and scored 75% or above on the test and just begun their training in aerosciences (spent 3 weeks or less); or cadets who were waiting to begin their aerosciences training but scored 75% or above in the E.C.L.

(b) Cadets who have had their English ability assessed by the E.C.L. and were still scoring below 75% but spent six months or more doing English in the Academy in addition to their previous language background at schools.

All tests with the exception of the oral interview were administered by the writer with the co-operation of some colleagues in the English Department at the K.F.A.A. The oral interview was administered by three native British/English speakers who also work in the English Department at the Academy. Reasons for choosing native speakers have been illustrated earlier in chapter five.

Within the Academy resources the administration and scoring of the test Battery proved to be feasible. Since the tests were administered on several occasions, standard forms of administration have been maintained as far as possible throughout. In addition to the Test Battery an Academic English skills questionnaire was distributed among all available instructors of the Academy.

These instructors were:

(a) Flying Instructors,

(b) Aeroscience Instructors and

(c) English Language Instructors.

The questionnaire aimed at deciding which language skills (reading, listening, writing or speaking) were most important to the cadets' success in the Academy's classes, largely Aerosciences. The good response to the questionnaire, represented by additional detailed comments by instructors, might have been attributed to general concern about cadets' English proficiency. However, because the writer is well aware of the language skills needed by cadets to pursue their training in Aerosciences and flying training, results of the questionnaire were not utilized to collect data for what skills were to be included in the test Battery, instead they were used in the syllabus section (chapter two).

8.5 T.E.A.C. Internal Analysis

8.5.1. Item Analysis

This analysis was performed on the grammatical structure and the listening comprehension tests (lecturette) because they were based on independent items. The sample was examined for item facility and item discrimination. The resulting item validity and difficulty indices were employed together in order to determine which items were invalid.

A total of 15 items out of a maximum 86 were rejected, 11 items out of 73 were rejected from the grammatical structure test and 4 out of 13 were rejected from the listening comprehension test (lecturette). The rejected items had either an item facility or

discrimination index below +.25.

8.5.2 T.E.A.C. Reliability

Reliability coefficients for all sub-tests and its total were computed. The obtained reliability coefficients for the sample are presented below.

Reliability of T.E.A.C. sub-tests and total

Cloze	.84
Dictation 1	.86
Dictation 2	.87
Grammar	.72
Listening Comprehension	.42
T.E.A.C. total	.96

As for the oral interview, raters reliability, it was maintained that adherence to the content of the interview as well as to the scoring scale may in fact contribute to the reliability of the test and scoring.

8.6 Factor analysis

A factor analysis technique was performed. All variables (sub-tests) were entered in the computer for initial analysis. Principal Component Analysis extracted two factors, these factors were then rotated using Varimax rotation patterns. While with Principal Component technique the first factor loaded on every variable of the T.E.A.C. it did not seem as stronog when rotation was employed, neither did factor two seem as weak as it appeared in Principal Component Analysis. The result of the analysis of the rotation revealed the existence of two major factors which may suggest that in the T.E.A.C., the grammatical structure and the

cloze tests were aligned with one another. On the other hand the listening comprehension (lecturette), the dictation and the oral interview were aligned with one another. However, these two factors may be tentatively labelled reading comprehension versus listening comprehension. The main experimental finding of factor analysis as employed in this study was that the unitary competence hypothesis as claimed by Oller (1979, 1983) was not supported.

Both the factor analysis technique and the test inter-correlation were used as confirmation of the T.E.A.C.'s content and construct validity. The extraction of two factors and the moderate correlations between subtests give a slight evidence to the existence of these types of validity, i.e. content and construct.

8.7 ANOVA

The ANOVA technique was employed to test the hypothesis that there is a significant difference between levels of proficiency. Test of significance as employed supported this, and all the F values were highly significant for the individual tests as well as the T.E.A.C. total indicating that there is a significant difference in the overall level of proficiency between the different levels tested confirming that the two samples tested belonged to two different populations.

8.8 External Analysis

Two kinds of validity criteria, concurrent and predictive, were obtained for the T.E.A.C. For concurrent validity two criteria were employed, instructors' estimates of cadets' language proficiency (based on a 6-point scale). In addition the E.C.L. (despite the reservations about it, outlined earlier) was also used as a

criterion to investigate the T.E.A.C.'s concurrent validity; after all they are both tests aimed at language proficiency of non-native speakers of English.

For predictive validity, end-of-term exams' scores in 9 of the Aerosciences subjects taught in the Academy were obtained. These subjects are: Mathematics, Physics, Airmanship, Aeroengines, Aerodynamics, Flying Terminology, Meteorology, Navigation and Avionics. In addition, a tenth criterion was also used and this was results of averaging each cadet's scores in all the nine subjects.

Correlation coefficients between each of these criteria and the T.E.A.C.'s sub-tests and totals were obtained for almost all candidates who took part in the T.E.A.C.

The range of obtained significant coefficients for T.E.A.C.'s total for both criteria were +.57 to +.82 (concurrent/instructors' estimates) and +.52 and .70 with the E.C.L., for predictive, significant coefficients ranged from +.47 to +.69. It was also observed that some sub-tests acted as an additional virtue of T.E.A.C.'s validity for both concurrent and predictive, particularly in the case of predictive validity in which significant coefficient was as high as +.78.

8.9 Scaled scores

Conversion from raw scores to scaled scores for each test are presented in the appendix for all tests except the oral interview which was reported in terms of levels.

8.10 Cut off score

In the process of the T.E.A.C.'s validation studies, the writer experimented with two different cut offs, namely: 177, 194, in

percentages they are: 60% and 65%. The first conclusion drawn was that on the whole Aeroscience subjects did not need different cut offs. About 70% of the subjects were identified as successes when 60% was chosen as a cut off point. However when the cut off was raised to 65%, on the whole there was no significant difference; again about 70% were identified as successes in the criteria. However, the decision where one should draw the line may be based on two grounds. Firstly, what percentages of success of these candidates admitted is required and secondly the level of proficiency one requires for these adamitted candidates. For the purpose of the K.F.A.A. use at present 60%, i.e. a raw score of 177 (86/ss) seems appropriate. The reasons are:

Firstly, flying training is a very expensive and complex business. It involves (as it is done in the Academy) proficiency in English language + mechanical aptitude and knowledge of the subject matter. Rejecting cadets on the basis of a very high score, 70% for instance or above, in English, is a harsh decision and by this one may reject high potentials who eventually will make very good pilots.

Secondly, 60% on the T.E.A.C. provided a satisfactory figure around 70% successes.

Thirdly, we need more pilots.

It is, therefore, recommended that if the T.E.A.C. will be used as a placement test then cadets who score below 60% may in fact be given some remedial English.

It should be noted that results presented here, i.e. cut off scores, were based on predictive criteria rather than concurrent. Reasons are to do with the element of unreliability in so far as

obtaining the concurrent validity based on instructors' estimates. This is not of course to say that subject matter grades are totally valid, since they themselves are not solely linguistic criteria but involves so many other factors (discussed earlier), it is in fact the use of proficiency testing (T.E.A.C.) as employed here which dictates the criteria. In my opinion if predictive validity is to be investigated thoroughly and instances of failure due to language problems be isolated the following design (which is beyond the scope of this study) may be used:

Time 1	Time 2	Oral
Proficiency test in a language	Tests of subject matter The same questions but different media	Mcq Essay

8.11 Level of proficiency, an influencing variable

Analysis of variance (ANOVA) test of significance revealed that there was a significant difference between the two different levels of proficiency tested by the T.E.A.C. both within the individual tests as well as the overall language proficiency.

8.12 Conclusions

The T.E.A.C. was regarded as both a reliable and valid measure of the English proficiency of the cadets of the K.F.A.A. who are going to undergo training in Aerosciences and flying in Saudi Arabia within the limits of the samples tested. The development of the T.E.A.C. was based on previously determined principles; this may mean that the content validity of the test can be guaranteed. We knew exactly what we wanted to measure right before we started developing our test. The factor analysis gave us an indication as

to the shape of the construct of language proficiency as measured by the T.E.A.C. The presence of two factors does not support the unitary competence hypothesis but instead a divisibility of some kind. More important is that the concurrent and predictive validity were claimed for the T.E.A.C.

8.13 Hypotheses

In accordance with previous work on language proficiency testing for instance (Davies 1965, Alderson 1978, Moller 1982, Weir 1984) and in view of the existing literature on language testing, a number of hypotheses were set up. Each of the hypotheses had been examined in the light of evidence obtained from current developments in the field of language testing and from the data that have been obtained and already discussed and presented in the previous chapters. These discussions and the conclusions reached are presented below.

8.13.1 Hypothesis one

The first hypothesis was set up that the E.C.L.:

(a) has a negative back-wash effect over the teaching/learning situation,

(b) is not compatible with the current thinking in language testing,

(c) there was no positive relation between English proficiency as measured by the E.C.L. and the academic outcome.

These collective hypotheses were examined in the light of the available data obtained from the Academy's records, and also from the existing literature on languageg testing.

The E.C.L., it was considered, is not compatible with current thinking in language testing. Moreover no evidence was obtained as

with regard to the test predictive validity, i.e. there was no positive relation between the E.C.L. and the academic outcome. Coefficients ranged between [-0.068 and 0.11] N = 63.

8.13.2 Hypothesis two

The null hypothesis was set up that there was no positive relation between English proficiency as assessed by the T.E.A.C. on the one hand and instructors' estimates of their cadets' English proficiency, the E.C.L. and academic successes on the other. This hypothesis was examined in the light of the validity studies in chapter seven, and on the basis of the evidence found, the null hypothesis was rejected.

The T.E.A.C., is therefore regarded as a valid measure of English proficiency "for the cadets of King Faisal Air Academy within the sample tested, as assessed by cadets' instructors and the E.C.L..". It is also a valid predictor of academic success where English is used as a medium to teach Aerosciences in the Academy.

8.13.3 Hypothesis three

In view of claims in the literature made by writers in the field (Oller 1979-1983) that there exists one factor underlying all abilities in language proficiency, the null hypothesis was set up as:

'that language proficiency is unitary'.

This hypothesis was examined in tHe light of the data obtained. It was shown that the factor analysis technique employed to investigate this hypothesis is necessarily a three-step process. A principal component analysis, followed by a rotation technique of some kind (varimax, oblique, equamas, quartimax etc.) and finally correlations of the factors extracted. In this study we have demonstrated that a

researcher may infer wrongly the existence of one factor only, on the basis of Principal Component Analysis. This inference as indicated above is a jump to an early conclusion and that the analysis does not stop there but should be followed by other steps.

However, this study revealed that there exists more than one factor underlying the construct of language proficiency as measured by the T.E.A.C. The principal component analysis revealed that there is both a general factor (but not unitary) and a smaller additional factor. The rotation technique as employed gave us a different loading pattern. The factors extracted remained two but factors extracted may be labelled reading comprehension versus listening comprehension. It was therefore concluded that the unitary competence hypothesis is no longer supported and that there exists more than one factor underlying linguistic abilities as measured by the T.E.A.C.

8.13.4 Fourth hypothesis

The null hypothesis was set up that there was no significant difference between different levels of proficiency as measured by the T.E.A.C.

A test of significance (ANOVA) was employed and on the basis of evidence obtained, the null hypothesis was rejected indicating that the T.E.A.C. may be used as a placement test in addition to its prime consideration, i.e. determining proficiency in English for the cadets of K.F.A.A. It was therefore considered that the two levels tested, level 1 and level 2 belonged to two different populations.

8.14 Implications and the future

The implications to be discussed here are two fold. Firstly, pedagogic implications (for the language syllabus, the teaching materials etc.) and secondly, practical and theoretical implications as with regard to the results of the T.E.A.C. as a measure of language proficiency for a particular discipline.

8.14.1 T.E.A.C. as guide to the language curriculum at the K.F.A.A.

back-wash effect

In addition to the first aim of language testing which is research, one additional aim which is not less important is the influence of the language test over the teaching/learning situation.

The fact that testing is an integral part of the teaching/learning process is indisputable. In our particular case study we demonstrated earlier (chapter one) the influence of the E.C.L. over the teaching/learning situation.

In chapter two we implicitly isolated three stages in the teaching/learning situation which are closely related to the teaching/testing processes as well as to the psychological factor (motivation). In any situation where proficiency in a language is a prerequisite for subsequent training and where a language course is provided, three distinct testing programmes may be distinguished. Pre-test, in-test and a post-test. The pre-test in most cases acts as a placement which takes place prior to a language programme, its immediate aims are: Firstly to promote on the part of the learner the awareness of the teaching/learning goals and in addition it may serve to promote motivation (Rea 1985). The in-course test normally monitors the students' progress and may be used as a diagnostic tool

to have some feedback on teaching/learning situations, such as the effectiveness of particular teaching methodologies and/or problems encountered by learners. The final stage of testing is to do with achievement or the standard of proficiency required. The stages identified above show how the teaching/testing programmes are closely related.

Apart from its eventual use as a measure of language proficiency, the T.E.A.C.'s role in the first place is viewed as an evaluative instrument of the whole language programme. Firstly it was recognized from the outset that cadets' interest lies in the ability to understand, speak, read and to some extent write in English. English language as employed by the Academy's cadets is central to their activities in Aerosciences and flying. Language teaching should aim to develop the three skills which was shown to be of importance to them.

8.14.2 Pedagogic grammar

We argued earlier that effective communication in a second language may not be attained unless both the linguistic competence and the effective use of the language in actual communicative situations have both been maintained. If linguistic accuracy is not a prerequisite for communication in some disciplines, it certainly is in the case of flying training and can be a matter of life and death. This is why, apart from arguments to do with the notion of proficiency in language presented earlier, it was decided to include tests on some linguistic categories which are tests of accuracy and of the system of the language in the first place rather than tests of language use. Even with the remaining tests, "the cloze, the

oral interview and the dictation tests", accuracy of syntax, phonology were prime factors in the assessment of these texts.

The grammatical categories entered in the grammatical structure test needs to be taught through pedagogic grammar exercises. A pragmatic approach may be highly desirable since it might be the case that no one theory of teaching grammar is complete in its own right. Teachers, syllabus designers should in fact be so flexible that they may allow for a multiple line of approach (Allen 1974). I am not going to suggest any one approach regarding teaching grammar for the cadets of the K.F.A.A., but it should be taught.

8.15 Theoretical and practical implications

Any proficiency test which will eventually undergo frequent use needs at least one additional parallel form so that the one version does not expose itself to over-use that may lead to tests being unobjective over time. A parallel form, from a proficiency testing point of view, acts as a check on test stability (reliability) and test content and concurrent validity. Comparable forms of language test reinforce the claims made for validity of the first version. The T.E.A.C. now exists in one form only; it is therefore necessary that parallel forms be designed and validated together with the first form.

8.15.1 Progress and achievement tests within the proposed course

When the English language instructors were invited to comment on the T.E.A.C.'s content, some expressed a concern with regard to the T.E.A.C.'s level of difficulty and the fact that it was beyond the level of some cadets in the English language training programme. However this was appreciated and the test was not given but to those

cadets whom it was felt were potential candidates for the level of proficiency required by the T.E.A.C. and consequently the level of proficiency required by subsequent training. While instructors agreed that the level of proficiency introduced in the T.E.A.C. is the level desired, they wanted their cadets at various levels of proficiency in the English Language Training programme (ELT) to be trained to take tests such as the cloze, the interview and the dictation. This again was appreciated and in order to account for the diversity at various levels of proficiency tests, the following may be useful guidelines to training at taking these tests. For cloze, training could/should start with simple sentences in which some words are missing. One may start with basic functional words, then the level, the length and complexity of style, structure and text may increase with the level of proficiency. In general, training over the gap filling technique is required to do the cloze test. So far as the oral interview is concerned, classes devoted to oral communication are necessary. Structured interviews at various levels of difficulty, group discussions on various topics of interest to cadets may be introduced. Controlled and free conversations may also be introduced at higher levels of proficiency. Training to take dictation may begin with simple sentences. The speed, the length, the complexity of style may be increased as the level of proficiency increased.

In general the input for the teaching material should draw on these appoaches encouraging towards the use of language in these skills included in the T.E.A.C.

8.16 Use of T.E.A.C.

In addition to its use as an evaluative instrument of the language programme at the K.F.A.A., it is hoped, having established that it is both a reliable and a valid measure of cadets' language proficiency within the limitations of the sample tested, that it will be in operation soon in place of the E.C.L. as part of the selection process for cadets to proceed to aerosciences and flying training in Saudi Arabia. If, on the other hand, comparable institutions would wish to make use of the T.E.A.C., the writer would not object provided the R.S.A.F. agrees and the test security and control will be in good hands.

8.17 Theoretical implications in relation to practical (teaching)

Factor analysis

The results of the factor analysis as employed here and elsewhere, Palmer and Bachman (1981), Farhady (1983), Vollmer and Sang (1983) indicated that the atomistic approach to language teaching is still as valid i.e. language skills can be taught and tested separately. Therefore, when one particular skill is emphasised, it is likely that cadets do better on it rather than a less emphasised one.

8.17.1 T.E.A.C.'s input material

It was assumed, right from the outset, that cadets may do better on test materials relevant to what they do. Not only they may do better but motivation may be preserved. Although I did not experiment with this to observe the extent to which cadets who are familiar with the tests' texts may appear to be more advantageous than students who are not (or less) familiar with the texts. Recent studies, Alderson and Urquhart (1983) and Erickson and Molloy (1983)

indicated that students did better on texts taken from their own particular discipline than those who were less familiar with the texts. The inclusion, therefore, of materials relevant to cadets in the T.E.A.C., it was assumed, they may do better on them. This may indirectly raise some theoretical issues. Among them that language may in fact be divided into varieties and registers for specific disciplines. It, therefore, follows that with this theoretical not experimental implication, common sense may suggest that the unitary competence hypothesis may no longer be supported and that language is indeed more complex.

8.18 Validity problems

8.18.1 Test content

We argued earlier that the universe to sample from the input material for this test and indeed for any proficiency test is so vast, a complete content validity for a proficiency test may never be fully established or even claimed. The T.E.A.C. directly tests three language skills, reading, listening and speaking. Although the dictation test involves an element of a writing task, its prime consideration is to test listening ability rather than creative writing, guided or not. Such a test from the point of view of language proficiency or better a proficiency in a language is incomplete on the basis of omission of one important skill. However, proficiency in terms of the language skills as defined in this context, i.e. the Academy's context is geared to competence in three skills, viz. listening, reading and speaking. Writing, it was shown, was the least emphasized skill. In taking Aerosciences examinations, cadets are only required to answer multiple choice

subject matter questions, writing down some brief definitions or else filling in gaps with appropriate specialized words.

However, scoring writing and speaking tests have been shown to be both reliable and feasible (E.L.T.S. 1978). So if particular situations need these tests to be included I cannot see how they can be omitted.

8.18.2 Predictive validity

The T.E.A.C. was shown to be a valid test in terms of its predictive power. Significant moderate to high correlation coefficents between the T.E.A.C. and the external criteria used to determine this type of validity were obtained. The use of end of term exam marks in a subject matter which involves, in addition to the linguistic factor many other factors, may be seriously questioned. The extent to which subject matter markers gave penalties for inappropriate use of language in answering questions is a factor to be considered. It might be the case that in order to determine the criterion-related validity of a proficiency test, the criterion needs to be a purely linguistic one, since it follows from this that the prime consideration of a proficiency test, as has been argued earlier, is one to do not with eventual success or failure in a subject matter as such, but the adequacy of one's language to follow a particular subject matter taught in the medium of that language.

This necessarily needs a longitudinal study and now it has been shown that T.E.A.C. has a satisfactory predictive validity what is needed is a follow-up study.

References and work consulted

Aitken, K.G. (1977)
Using cloze procedure as an overall language proficiency test. TESOL Quarterly, 11, 1, pp.59-67.

Al-Farsy, F. (1978)
Saudi Arabia a case study in developments. London Staty International.

Alderson, J.C. (1978)
A study of the cloze procedure with native and non-native speakers of English. University of Edinburgh. Unpublished Ph.D. thesis.

_____ (1979)
The cloze procedure and proficiency in English as a foreign language in TESOL Quarterly vol. 13. June no.2.

_____ (1981)
Report of the Discussion on Testing English for specific purposes. In III Issues in Language Testing. ELT documents. The British Council. Alderson, J.C. and Huhges, A. (eds.).

_____ (1981)
Report of the discussion on general language proficiency in III Issues in Language Testing. Alderson, J.C. and Hughes, A. (eds.). The British Council.

Alderson, J.C. (1986)
Innovations in language testing? In Innovations in language testing (ed.) Matthew Portal (1986). Wheaton and Co. Ltd., Exeter.

Alderson, J.C. and Hughes, A. (eds.) (1981)
ELT Documents. III Issues in Language Testing. The British Council.

Alderson, J.C. and Urquhart, A.H. (1983)
The effect of student background discipline on comprehension, a pilot study. In Hughes, A. and Porter, D., editors. Current developments in language testing. London: Academic Press.

Allen, J.P.B. and Davies, A. (eds.) (1977)
Testing and experimental methods. Edinburgh Course in Applied Linguistics Vol. 4, Oxford: Oxford University Press.

Allen, J.P.B. and Widdowson, H.G. (1975)
Grammar and language teaching. In Allen, J.P.B. and Corder, S.P. eds. The Edinburgh Course in Applied Linguistics, vol. 2, London: Oxford University Press.

American Language Course Volumes 400 - 2600
Defense Language Institute, Lackland Air Force Base. United States of America.

American Language Course Placement Test.
Defense Language Institute, Lackland Air Force Base. United States of America. ALCPT Form IIR.

Anastasi, A. (1961)
Psychological testing. Second edition London. Macmillan.

Anderson, J. (1972)
A technique for measuring comprehension and readability. English Language Teaching 25 pp.178-82.

_____ (1972)
The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. English Language Teaching, 27, 1, pp.66-72

Bakalla, M.H. (1984)
Arabic culture through its language and literature. Kegan Paul International.

Bormouth, J.R. (1965)
Validities of grammatical and semantic classifications of cloze test scores. In Figuerel, 283-285.

_____ (1967)
Comparable cloze and multiple-choice comprehension test scores. Journal of Reading 10: 291-299.

Bowen, J.D. (1976) Current research on an integrative test of English grammar. RELC Journal 7: 30-37

Brown, J.D. (1983)
A closer look at cloze validity and reliability. In issues in Language Testing Research (ed.) Oller, J.W., Jr. Newbury House Publishers, Inc. Rowley, Massachusetts.

Brumfit, C.J. and Johnson, K. (eds.) (1979)
The communicative approach to language teaching. Oxford University Press.

Campbell, D.T. and Fiske, D.. (1959)
Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56: 81-105.

Canale, M, and Swain, M. (1980)
Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics 1, 1980: 1-47.

_____ (1981)
A Theoretical Framework for Communicative Competence in: The Construct Validation of Tests of Communicative Competence (eds.) Palmer, A.S., Groot, P.J.M. and Trosper, G.A. Teachers of English to speakers of other languages, Washington D.C., U.S.A.

Carroll, B.J. (1978)
English Language Testing Service. Specifications in British Council ELT Documents, Issues in language testing (eds.) Alderson. J.C. and Hughes, A.

_____ (1980)
Testing Communicative Performance. Oxford: Pergamon Press.

_____ (1983)
The English Language Testing Service in "Case Studies in ELT". ed. Jordan, R.A. William Collins Sons and Co.

Carroll, J.B. (1961)
Fundamental considerations in testing for English Language Proficiency of foreign students. In Center for Applied Linguistics. Testing the English proficiency of foreign students. Washington D.C.: Author. Reprinted in Allen and Campbell, 1972. 313-320.

_____ (1962)
The prediction of success in intensive foreign language training. In: Training Research and Education. University of Pittsburg Press.

_____ (1973)
Implications of aptitude test research and psycholinguistic theory for foreign language teaching. Linguistics 112: 5-14.

_____ (1968)
The psychology of language testing. In Davies, A. ed. (1968).

_____ (1983)
Psychometric theory and language testing. In: Issues in Language Testing Research (ed. Oller, J.W., J.R. (1983). Newbury House.

Carroll, J.B. and Sapon, S.M. (1959)
The Modern Language Aptitude Test. New York: The Psychological Corporation.

Chomsky, N. (1965)
Aspects of the theory of syntax. The M.I.T. Press.

Clifford, R. (1980)
In: Measuring spoken language proficiency. Frith, J. (ed.). Washington University Press.
Foreign Service Institute Factor Scores and Global Ratings.

Cohen, A.D. (1980)
Testing language ability in the classroom. Rowley, Mass. Newbury House.

Condrad, C.A. (1970)
. The cloze procedure as a measure of English proficiency. Unpublished M.A. thesis UCLA.

Crawford, A. (1970)
The cloze procedure as a Measure of Reading Comprehension of Elementary level Mexican-American and Anglo-American Children. Unpublished Ph.D. dissertation, University of California, Los Angeles.

Criper, C. (1981)
. Reaction to the Carroll paper. In: Issues in Language Testing. ELT documents. The British Council. Eds. Charles Alderson and Arthur Hughes.

Cronbach, L.J. (1960)
Essentials of Psychological Testing 2nd edition. New York. Harper and Row.

_____ (1971)
Test Validation in R.L. Thorndike, ed. Educational Measurement. 2nd ed. Washington D.C. American Council on Education.

Cummins, J. (1979)
'Cognitive/academic language proficiency, linguistic interdependence, the optimal age question and some other matters'. Working papers on Bilingualism 19: 197-205.

_____ (1980)
'The cross-lingual dimensions of language proficiency: implications for bilingual education and the optimal age issue'. TESOL Quarterly 14: 175-187

Darnell, D.K. (1976)
Clozentropy: a proceudre for testing English language proficiency of foreign students. Speech Monographs 37, 1, pp.36-46.

Davies, A. (1964)
English Proficiency Test Battery, Version A. London. British Council.

_____ (1965)
Proficiency in English as a second language. Unpublished Ph.D thesis. University of Birmingham.

_____ (1965a)
English Proficiency Test Battery, short version. form B. London. British Council.

_____ (1968) (ed.)
Language Testing Symposium: A Psycholinguistic Approach. London: Oxford University Press.

_____ (1977)
The construction of language tests. In: Testing and Experimental Methods. J.P.B. Allen and Alan Davies, (eds.).

_____ (1978)
Language testing. In: Language teaching and linguistic abstracts, vol. 11, nos. 3/4, 1978 pp 145-160 and 215-232.

_____ (1981)
A Review of Communicative Syllabus Design by J. Munby, in TESOL Quarterly, vol. 15/3 pp. 332-336.

_____ (1984)
Validating three tests of English language proficiency in: Language testing 1/1 1984. Edward Arnold pp. 50-69.

_____ (1985)
Follow my leader: Is that what language tests do.
In: New Directions in Language Testing Research.

_____ (1986)
Indirect ESP Testing: Old Innovations in: Innovations in language testing (ed.) Matthew Portal (1986)
Wheaton and Co. Ltd., Exeter.

Duran, R.P. (1984)
Some implications of communicative competence research for integrative proficiency testing. In: Communicative Competence Approaches to Language Proficiency Assessment: Research and Application. ed. Rivera Charlene.
Multilingual Matters Ltd.

Erickson, M. and Molloy, J. (1983)
ESP test development for engineering students. In Issues in language testing research. Ed. Oller:
Newbury House.

Farhady, H. (1983)
On the plausibility of the Unitary language proficiency factor in: Issues in language testing research (ed.) Oller, J.W., J.R. Newbury House.

_____ (1983)
The dijunctive fallacy between discrete-point and integrative tests in: Issues in language testing research. (ed.) Oller, J.W. J.R. Newbury House.

Ferguson, A. (1959)
"Diglossia" Word no. 15. Reprinted in D. Hymes (ed.)

Fountain, R. (1974)
A case for the dictation test in selection of foreign students for English medium study in New Zealand.
University of Edinburgh, Department of Linguistics.

Frith, J.R. (ed.) 1980
Measuring spoken language proficiency. Washington D.C. Georgetown University Press.

Gallant, R. (1965)
Use of cloze tests as a measure of readability in the primary grades. In Figuerel, 286 2 87

Gardner, R.C. and Lambert, W.E. (1972)
Attitudes and Motivation in second language learning.
Rowley, Mass. Newbury House,

Greenberg, K.L. (1986)
The development and validation of the TOEFL Writing Test: A discussion of T.O.E.F.L. Research Reports 15 and 19. TESOL Quarterly vol. 20, no. 3, 1986.

Guilford, J.P. and Fruchter, B. (1973)
Fundamental statistics in psychology and eduction. Tokyo: McGraw-Hill

Hafner, L.E. (1964)
Relationship of various measures to the cloze. Thirteenth year book of the National Reading Conference, 135-145.

Halliday, M., Mcintosh, A., Strevens, P. (1964)
The linguistic sciences and language teaching. London. Longman.

Halliday, M.A.K. (1970)
Language structure and language function. In J. Lyons, ed. New Horizons in Linguistics. Harmondsworth, England: Penguin Books.

Harris, D.P. (1969)
Testing English as a second language. New York, McGraw-Hill.

Haskell, J.F. (1973)
Refining the cloze testing and scoring procedures for use with ESL students. Columbia University, Ed. D. Thesis.

Harrisson, A. (1983)
A language testing handbook MacMillan Press, London.

Heaton, J.B. (1975)
Writing English Language Tests. London: Longman.

Heaton, J.B. and Pugh, A.K. (1974)
A study of the relationship between scores obtained by overseas students on a Test of English proficiency and the examination results in their University course. University of Leeds: School of Education.

Howatt, A.P.R. (1974)
The Background to course design in Allen, J.P. and Corder, S.P (eds.). The Edinburgh Course in Applied Linguistics vol. 3. London. Oxford University Press.

Howatt, A.P.R. and Davies (1979)
ELBA testing 1973-77. Interim Report on analysis of results. Edinburgh University: Department of Linguistics.

Hudson, R. (1980)
Sociolinguistics. Cambridge University Press.

Hughes, A. (1986)
A pragmatic approach to criterion - referenced foreign language testing. In: Innovations in language testing (ed.) Matthew Portal (1986).
Wheaton and Co. Ltd., Exeter.

Hutchinson, T. and Waters, A. (1980)
Communication in the technical classroom: 'You just shove this little chappie in here like that' in: Projects Materials Design. ELT documents special.
The British Council.

Hymes, D. (1967)
Models of the interaction of language and social setting. Journal of social issues 23, 2:8-28

_____ (1968)
The ethonography of speaking. In J. Fishman, ed.
Readings in the sociology of language. The Hague: Mouton.

_____ (1972)
'On communicative competence' in Pride and Holmes 1972.

Ingram, E. (1964)
English Language Battery (ELBA). University of Edinburgh, Department of Linguistics.

_____ (1973)
English standards for foreign students: University of Edinburgh Bulletin 9, 12, 4-5.

Irvine, P., Atai, P. and Oller, J.W. (1974)
Cloze, dictation and the Test of English as a foreign language. Language Learning. 24, 2, 245-52.

Jakobovits, La. (1970)
Foreign Lanugage Learning (Rowley, Mass: Newbury House).

Jones, R.L. and Spolsky, B. (eds.) (1975)
Testing Language Proficiency. Washington D.C.: Center for Applied Linguistics.

Jongsma, E. (1971)
The cloze procedure as a teaching technique. Blooming, Ind.: Indiana University.

Johnson, K. (1982)
Communicative Syllabus Design and Methodology. Oxford: Pergamon Press.

Klein-Braley, C. (1983)
A cloze is a cloze is a question. In: Issues In Language Testing Research (ed.) Oller, J.W. Newbury House.

Lado, R. (1957)
Linguistics Across Cultures. Ann Arbor, Michigan. University of Michigan.

_____ (1961)
Language Testing. New York: McGraw-Hill and London: Longman.

Lee, Y.P., Fok, C.Y.Y., Lord, R. and Low, G. (1985)
New Directions in Language Testing. Pergamon Institute of English.

Lockheed, English Language Training Program (1981)
A Word Count and Contrastive Frequency Level Rating of the American Language Course.

Mackay, R. and Mountford, A. (eds.) (1978)
English for Specific Purposes. London: Longman.

Madsen, H.S. and Jones, R.L. (1981)
Classification of Oral Proficiency Tests. In: The Construct Validation of Tests of Communicative Competence. Eds. Palmer, Groot and Trosper. TESOL Washington.

Mason, V. (1972)
Report on cloze tests administered to Thai students. Bulletin of the English Language Centre, 2, 1, 31-9.

Maxwell, A. (1965)
A comparison of two English as a foreign language tests. University of California (Davis).

Mcdonough, J. (1984)
English for Specific Purposes

Moller, A.D. (1981)
Assessing proficiency in English for use in further study. In J.A.S. read (ed.) Directions in Language Testing pp.58-71. Singapore: Singapore University Press/SEAMEO Regional Language Centre.

_____ (1981)
Reaction to the Morrow papoer in Issues of Language Testing. III. ELT documents. The British Council.

_____ (1982)
A study in the validation of proficiency tests of English as a foreign language. Unpublished Ph.D thesis. University of Edinburgh. Department of Linguistics.

Morrow, K.E. (1982)
Techniques of evaluation for a notional syllabys: Reading: Centre for Applied Language Studies. University of Reading. (Study commissioned by the Royal Society of Arts).

_____ (1981)
Communicative Language Testing. Revolution or Evolution, in III Issues in Language Testing. ELT documents. The British Council. London.

Munby, J. (1978)
Communicative Syllabus Design. Cambridge, Cambridge University Press.

Nemser, W. (1974)
Approximative Systems of Foreign Language Learners. In: Error analysis: Perspectives on Second Language Acquisition (ed.) Richards, J., Longman.

Oller, J.W. (1971)
Dictation as a device for testing foreign language proficiency. English Language Testing 25, 3, 254-9.

_____ (1972)
Cloze tests of second language proficiency and what they measure. Language Learning 23, 1, 105-18.

Oller, J.W. Jr. (1979)
Language TESTS at school. Longman.

_____ (ed.) (1983)
Issues in Language Testing Research. Rowley. Mass. Newbury House.

_____ (1983)
Evidence for a general language proficiency factor: an expectancy grammar. In Issues in Language Testing Research. Rowley. Mass. Newbury House.

_____ (1972c)
Scoring methods and difficulty levels for tests of proficiency in English as a second language. Modern language journal 56: 151-158.

_____ (1973)
Cloze tests of second language proficiency

Oller, J.W. and Conrad, C. (1971)
The cloze technique and ESL proficiency. Language Learning 21: 183-195.

Oller, J.W. and Streiff, V. (1975)
Dictation: a test of grammar-based expectancies. English Language Teaching Journal,30, 1, 25-36.

Oller, J.W. and Perkins, K. (1980)
Research in language testing. Rowley, Massachusetts: Newbury House.

Palmer, A.S. and Bachman, L.F. (1981)
Basic concerns in test validation. In: Issues in Language Testing. ELT documents. The British Council.

Palmer, A.S. and Groot, J.M. (1981)
An introduction in: The Construct Validation of Tests of Communicative Competence. T.E.S.L. Washington D.C.

Palmer, A.S., Groot, J.M. and Trosper, G.A. (1981) eds.
The Constant Validation of Tests of Communicative Competence T.E.S.L. Washington D.C.

Petersen, C.R., Cartier, F.A. (1975)
Some theoretical problems and practical solutions in proficiency test validity. In: Testing Language Proficiency, Jones, R. and Spolsky, B. (eds.) Center for Applied Linguistics, Arlington, Virginia.

Pike, L.W. (1973)
An evaluation of present and alternative item formats for use in the Test of English as a foreign language. Unpublished manuscript, Educational Testing Service, Princeton, New Jersey.

Potter, T.C. (1968)
A taxonomy of cloze research. Technical Report No. 1. Inglewood, Calif.: Southwest Regional Laboratory for Educational Research and Development.

Rea, Pauline M. (1985)
Language Testing and the communcative language teaching curriculum. In New Directions in Language Testing. Lee, Y.P., Fok, A., Lord, R., Low, G. (eds.). Pergamon Press. Oxford.

Richterich, R. and Chancerel, J.L. (1980)
Identifying the Needs of Adults Learning a Foreign Language: Oxford. Pergamon Press.

Rivers, Wilga (1964)
The psychologist and the foreign language teacher. Chicago and London: University of Chicago Press.

_____ (1968)
Teaching foreign language skills. Chicago. University of Chicago Press.

Robinson, P. (1973)
Oral Expression Tests. English Language Teaching, vol. 25, nos. 2-2.

Ruddell, R.B. (1964)
A study of the cloze comprehension technique in relation to structurally controlled reading material. Improvement of Reading Through Classroom Practice: 9:298-303.

Schulz, R.A. (1977)
Discrete point versus simulated communication testing in foreign languages. Modern language journal 61, 3, 91-101.

Shohamy, E. (1983)
Interrater and intrarater reliability of the oral interview and concurrent validity with cloze procedure in Hebrew. In Issues in Language Testing Research. Oller. J.W., J.R. ed. Newbury House.

Scott, M.S. and Tucker, G.R. (1974)
Error Analysis and English Language Strategies of Arab Students. In Language Learning 1974 pp.69-96.

Spolsky, B. (1975, 1978)
Language testing: art or science. Paper originally presented at the fourth international congress of Applied Linguistics, published in vol. 3 Stuttgart in 1978: Language Testing. ed. Nickel, G. Hochschulverlag.

_____ ed. (1978)
Advances in Language Testing: Series 2, Approaches to Language Testing. Arlington Virginia: Centre for Applied Linguistics.

_____ ed. (1979)
Advances in Language Testing: Series 1, Some Major Tests. Arlington Virginia: Centre for Applied Linguistics.

Stern, H.H. (1983)
Fundamental Concepts in Language Teaching O.U.P.

Streiff, V. (1978)
Relationships among oral and written cloze scores and achievement test scores in a bilingual setting. In Oller and Perkins: Language in education: testing the tests. Strevens, P. (1980) "English for special purposes: an analysis and survey in Croft, Readings in English as a second language pp. 458-472. 2nd Edition. Cambridge, Winthrop Inc.

Stubbs, J.B., Tucker, G.R. (1974)
The cloze test as a medium of ESL proficiency for Arab students in language learning.

Taylor, W.L. (1953)
Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 30, 415-33.

_____ (1956)
Recent developments in the use of the cloze procedure. Journalism Quarterly, 33, 42-9.

Test of English as a foreign language. Educational Testing Service. Princeton, New Jersey.

Valette, R.M. (1967)
Modern Language Testing: a handbook. New York: Harcourt Brace Jovanovich.

_____ (1977)
Modern Language Testing 2nd edition. New York: Harcourt Brace Jovanovich.

Van Ek, J. (1975)
The Threshold Level in a European Unit/ Credit System for Modern Language Learning by Adults. Strasbourg. 1975.

Van Eks, T. et al. (1984)
Applied Linguistics and the learning and teaching of foreign languages. Walters-Noordhoff.

Vollmer, H. (1981)
Why are we interested in General Language Proficiency? In III Issues in Language Testing. ELT documents. (eds.) Alderson, J.C. and Hughes, A. The British Council.

Vollmer H. and Sang, F. (1983)
Competing hypotheses about second language ability: a plea for caution in Issues in Language Testing Research. (ed.) Oller, J.W., J.R. Newbury House.

Weir, Cyril. (1981)
Reaction to the Morrow Paper in Issues in Language Testing. The British Council. ELT documents III.

_____ (1983)
Identifying the language problems of overseas students in tertiary education in the United Kingdom. University of London, unpublished Ph.D. thesis.

Widdowson, H.G. (1978)
Teaching Language as Communication. London: Oxford University Press.

_____ (1983)
Learning Purposes and Language Use. London: Oxford University Press.

Wilkins, D.A. (1976)
Notional Syllabuses. Oxford: O.U.P.

Willis, D. (1981)
Needs Analysis and Syllabus Specification. In English Language Research Journal, No. 2. 1981

Yalden, J. (1983)
The Communicative Syllabus: Evolution, Design and Implementation. Pergamon Press.

Zhong, Y. and Chun, G.S. (1985)
The English Proficiency Test Used in China: A Report.
In: New Directions in Language Testing (eds.)
Lee, Y.P., Fok, C.Y.Y., Lord, R. and Low, G. Pergamon
Press.

TABLE OF CONTENTS

Appendices

LIST OF TABLES IN THE APPENDIX

TEST OF ENGLISH FOR AIR CADETS

PILOT STUDY

Test 1

Name: Date:

Service Number:

Course:

Instructions

This is a test of how well you can recognize correct English grammar. Each item in this test is a short sentence or a short conversation with one or more words missing. Choose the ONE answer that is grammatically correct.

Example: I _____ to school yesterday.

a) go

b) went

c) am going

d) had gone

For this sentence, went is grammatically correct. It is answer b, so if this were part of the test, you would circle b on your booklet.

(1) My father is _____ in the Air Force.

a) fighter pilot

b) a fighter pilot

c) the fighter pilot

d) this fighter pilot

(2) your aircraft needs _____

a) checking

b) to check

c) check

d) checked

(3) The pilot was in contact with the tower _____ the entire flight.

a) since

b) while

c) across

d) during

(4) "_____ are you flying solo?

"Next month"

a) why

b) when

c) where

d) how

(5) Water _____ of hydrogen and oxygen.

a) consisting

b) has consisted

c) is consisted

d) consists

(6) He _____ in Riyadh a few days ago.

a) is arriving

b) arrives

c) arrived

d) has arrived

(7) Are those the Cadets _____?

a) you were talking about them

b) you were talking about

c) about that you were talking

d) that you were talking about them

(8) While I was reading, the telephone _____.

a) rung

b) ringing

c) rang

d) had rung

(9) He enjoys _____ jet planes.

a) flies

b) flying

c) to fly

d) to flying

(10) We saw two aircraft _____ the runway.

a) into

b) at

c) on

d) in

(11) I have been reading this week's aviation journal now _____.

a) two hours ago

b) since two hours

c) before two hours

d) for two hours

(12) "Have you finished your research exercise?"

"Yes, it _____ now".

a) typed

b) is being typed

c) types

d) is typing

(13) Airframes are normally made _____ light steel.

a) of

b) by

c) off

d) for

(14) "_____ does your father go to Jeddah on business?

"At least once a week"

a) How often

b) How many

c) How long

d) When

(15) We _____ to work hard or we will not pass the test.

a) should

b) have

c) must

d) had

(16) Captain Wilkins has gone to France _____ holiday.

a) for

b) by

c) on

d) in

(17) Neither Hassan _____ Fahad had taken an English Course before joining the Academy.

a) either

b) and

c) nor

d) or

(18) You are not a cadet in this Academy _____?

a) are you

b) aren't you

c) were you

d) weren't you

(19) Don't you enjoy _____ by plane?

a) travelling

b) to travelling

c) to travel

d) travel

(20) The atmosphere is a mixture _____ gases.

a) in

b) of

c) off

d) with

(21) Pilots _____ fly their aircraft without checking them first.

a) ought not

b) must not

c) have not

d) do not have

(22) Have you flown the aircraft _____ has had its engin repaired?

a) whose

b) which

c) of which

d) of whom

(23) My brother _____ join the K.F.A.A. but he is not very sure yet.

a) can

b) would

c) may

d) will

(24) _____ studied English before joining the Academy, hadn't you?

a) You had

b) Had you

c) Have you

d) Hadn't you

(25) Captain Ali wanted to know _____

a) was that my uniform

b) is that my uniform

c) if that was my uniform

d) that was my uniform

(26) The cockpit is _____ that it cannot accommodate two pilots.

a) too small

b) so small

c) very small

d) small enough

(27) A normal landing _____ with the landing gear up.

a) not is made

b) not made

c) is made not

d) is not made

(28) He is _____ in the Army.

a) an officer

b) officer

c) a officer

d) the officer

(29) Aircraft pilots communicate _____ control towers.

a) at

b) for

c) during

d) with

(30) "Please turn the television down. I _____ to sleep".

a) am trying

b) tried

c) try

d) have trying

(31) We _____ the flying wing two months ago.

a) visited

b) have visited

c) were visited

d) visit

(32) I am sure Meteorology is not _____ as Aerodynamics.

a) more difficult

b) as difficult

c) so difficult

d) very difficult

(33) _____ aircraft on this ramp have served the Air Force for a long time.

a) a little

b) every

c) all

d) a few

(34) The wings are _____ to the fuselage.

a) attaching

b) attached

c) attaches

d) attach

(35) _____ it was necessary to attend the lecture on Aeroengine, neither Ali nor Abdulazaz did.

a) However

b) But

c) Although

d) Nevertheless

(36) The aircraft was mistaken _____ an eagle.

a) to

b) by

c) with

d) for

(37) I wish I _____ how to fly that aircraft.

a) knew

b) know

c) will know

d) would know

(38) Is that the runway _____ the aircraft stopped?

a) which

b) where

c) when

d) were

(39) He _____ swimming regularly, but now he never goes.

a) was used to go

b) used to go

c) use to go

d) goes

(40) He was the best pilot _____ them all.

a) of

b) from

c) by

d) than

(41) He has been studying English for six months now and _____.

a) still he can't read properly

b) he still can't read properly

c) he can't still read properly

d) he can still not read properly

(42) The pilot _____ life was saved is my best friend.

a) whose

b) of whom

c) when

d) of which

(43) Did you see the photographs _____ at the graduation?

a) I had taken

b) I had taken them

c) had taken

d) that I had taken them

(44) He left for Jeddah _____ Tuesday.

a) at

b) on

c) by

d) in

(45) I have read the book _____ you lent me.

a) that

b) whom

c) whose

d) what

(46) If he had flown at a lower level he _____ avoided the birdstrike.

a) could had

b) would have

c) will have

d) should have

(47) If I _____ a student in the University I would read for a degree in Physics.

a) am

b) were

c) was

d) had been

(48) When the wing stalls there is a _____ in lift and a increase in drag.

a) reduction

b) reduced

c) reduce

d) reducing

(49) As the pilot _____ the runway, he saw an aeroplane taking off.

a) approach

b) was approaching

c) is approaching

d) approaching

(50) "Have you flown solo?" I asked my roommate.

I asked my roommate _____

a) had he flown solo

b) if he had flown solo

c) that he had flown solo

d) had you flown solo

(51) Did you meet the pilot _____ wife gave birth to a baby boy.

a) whose

b) that

c) whom

d) which

(52) If the air hadn't been turbulent the pilot _____ to climb.

a) wouldn't had decided

b) wouldn't decided

c) wouldn't have decided

d) wouldn't has decided

(53) The temperature of air usually decreases _____ altitude.

a) in

b) by

c) into

d) with

(54) "By the end of next month _____ solo".

a) we will be flying

b) we would fly

c) we have flown

d) we fly

(55) "Where is your car?"

"It _____ in the garage"

a) is being repaired

b) is repaired

c) is repairing

d) is been repaired

(56) The Strikemaster is longer than the Cessna _____ five feet.

a) by

b) in

c) with

d) of

(57) This is _____ interesting book I have ever read.

a) most

b) the most

c) a most

d) the more

(58) "Are there many students in the lecture hall?"

"No, only _____ turned up."

a) a few

b) less

c) little

d) a little

(59) _____ you are sick you must stay in bed.

a) Although

b) Therefore

c) However

d) Because

(60) Yesterday we watched a film _____ flying.

a) about

b) around

c) of

d) on

(61) We _____ of the examination results two hours ago.

a) had informed

b) have informed

c) are informed

d) were informed

(62) Air traffic control officers give aircraft regular up-dates of weather condiditions _____ the airport.

a) in

b) on

c) at

d) about

(63) You _____ get your hair cut before the officer sees you.

a) should rather

b) could

c) should

d) may

(64) Water boils _____ you heat it to 100C°.

a) unless

b) if

c) until

d) although

(65) If he had worked hard he _____ passsed the examination.

a) would have

b) could have

c) will have

d) had had

(66) I wish you _____ it last week

a) could have sent

b) sent

c) could sent

d) have sent

(67) "He speaks very good English"

"Yes he does. He _____ in England for three years.

a) was learning the language

b) has been learning the language

c) has learned the language

d) have been learning the language

(68) Choose the correct sentence.

a) A first plane I flew was Cessna 172.

b) The first plane I flew was a Cessna 172.

c) A first plane I flew was a Cessna 172.

d) The first plane I flew was Cessna 172.

(69) "Is this your log book?"

"Yes, this logbook is _____"

a) mine

b) yours

c) my

d) your

(70) "Where have you been? We have been waiting for you _____ 12 o'clock"

a) for

b) during

c) between

d) since

(71) This palmtree _____ in 1982.

a) planted

b) plants

c) is planted

d) was planted

(72) The instructor took us to see the newly _____ plane

a) manufacture

b) manufacturing

c) manufactured

d) manufacturer

(73) Would you be interested in _____ a lecture on Mathematics?

a) attended

b) attending

c) to attend

d) attend

LISTENING COMPREHENSION TEST

(PILOT STUDY)

(1) Meteorology is one of the earths' sciences and is a study about:

a) the earth's surface

b) the oceans and the waves

c) the many regional and different climates

d) the clouds and the weather

(2) Meteorology affects:

a) only a few people

b) almost everyone of us

c) only military and civil pilots

d) only farmers, fishermen and pilots

(3) Turbulence causes an aircraft to:

a) land safely

b) be buffetted and thrown about the sky

c) be hidden by cloud

d) have a smooth flight

(4) Examples of severe weather conditions are:

a) icing and thunderstorm

b) strong squally winds and turbulence

c) tropical revolving storms

d) all of the above

(5) An aircraft flying across the Atlantic Ocean with a head wind of more than 100 knots is likely to:

a) arrive on time

b) arrive ahead of schedule

c) be about 2 hours late on arrival

d) be re-routed but still arrive on time

(6) A Jet Stream is:

a) an Easterly wind

b) a squally wind

c) a wind blowing at 100 knots or more

d) the wind at cruise level

(7) A weather forecast is:

a) the latest meteorological situation

b) a statement of what weather is expected to happen in the future

c) an accurate weather report

d) a reading of temperature, pressure and wind

(8) A red sunrise in the morning means that:

a) there's about a 70% chance of rain during that day

b) a red sunset will follow later in the day

c) it will soon be raining

d) the next day will be fine

(9) The World Meteorological Organisation is based at:

a) Frankfurt

b) Washington

c) Paris

d) Geneva

(10) A computer used for weather forecasting:

a) will always produce a correct weather forecast

b) relies on accurate information to be fed into it before the forecast is made

c) takes a long time to produce a forecast

d) relies on weatherlore for some of its information

(11) A satellite provides weather information about:

a) the wind

b) turbulence

c) the clouds

d) the visibility

(12) Observations about the weather are made:

a) every day of the year

b) only when required

c) when bad weather is forecast

d) only during the day and not at night

(13) The weather changes:

a) at regular intervals of time

b) always very slowly

c) only during the daytime

d) sometimes very quickly

TEST OF ENGLISH FOR

AIR CADETS

FINAL VERSION

TEST OF ENGLISH FOR AIR CADETS

Test 1

Name: Date:

Service Number:

Course:

Instructions

This is a test of how well you can recognize correct English grammar. Each item in this test is a short sentence or a short conversation with one or more words missing. Choose the ONE answer that is grammatically correct.

Example: I _____ to school yesterday.

a) go

b) went

c) am going

d) had gone

For this sentence, went is grammatically correct. It is answer b, so if this were part of the test, you would circle b on your booklet.

(1) My father is _____ in the Air Force.

a) fighter pilot

b) a fighter pilot

c) the fighter pilot

d) this fighter pilot

(2) your aircraft needs _____

a) checking

b) to check

c) check

d) checked

(3) The pilot was in contact with the tower _____ the entire flight.

a) since

b) while

c) across

d) during

(4) "_____ are you flying solo?

"Next month"

a) why

b) when

c) where

d) how

(5) Water _____ of hydrogen and oxygen.

a) consisting

b) has consisted

c) is consisted

d) consists

(6) He _____ in Riyadh a few days ago.

a) is arriving

b) arrives

c) arrived

d) has arrived

(7) While I was reading, the telephone _____.

a) rung

b) ringing

c) rang

d) had rung

(8) He enjoys _____ jet planes.

a) flies

b) flying

c) to fly

d) to flying

(9) We saw two aircraft _____ the runway.

a) into

b) at

c) on

d) in

(10) I have been reading this week's aviation journal now

_____.

a) two hours ago

b) since two hours

c) before two hours

d) for two hours

(11) "Have you finished your research exercise?"

"Yes, it _____ now".

a) typed

b) is being typed

c) types

d) is typing

(12) Airframes are normally made _____ aluminium.

a) of

b) by

c) off

d) for

(13) "_____ does your father go to Jeddah on business?

"At least once a week"

a) How often

b) How many

c) How long

d) When

(14) We _____ to work hard or we will not pass the test.

a) should

b) have

c) must

d) had

(15) Neither Hassan _____ Fahad had taken an English Course before joining the Academy.

a) either

b) and

c) nor

d) or

(16) You are not a cadet in this Academy _____?

a) are you

b) aren't you

c) were you

d) weren't you

(17) Don't you enjoy _____ by plane?

a) travelling

b) to travelling

c) to travel

d) travel

(18) The atmosphere is a mixture _____ gases.

a) in

b) of

c) off

d) with

(19) Pilots _____ fly their aircraft without checking them first.

a) ought not

b) must not

c) have not

d) do not have

(20) Have you flown the aircraft _____ has had its engine repaired?

a) whose

b) which

c) of which

d) of whom

(21) My brother _____ join the K.F.A.A. but he is not very sure yet.

a) can

b) would

c) may

d) will

(22) _____ studied English before joining the Academy, hadn't you?

a) You had

b) Had you

c) Have you

d) Hadn't you

(23) Captain Ali wanted to know _____

a) was that my uniform

b) is that my uniform

c) if that was my uniform

d) that was my uniform

(24) A normal landing _____ with the landing gear up.

a) not is made

b) not made

c) is made not

d) is not made

(25) He is _____ in the Army.

a) an officer

b) officer

c) a officer

d) the officer

(26) Aircraft pilots communicate _____ control towers.

a) at

b) for

c) during

d) with

(27) "Please turn the television down. I _____ to sleep".

a) am trying

b) tried

c) try

d) have trying

(28) We _____ the flying wing two months ago.

a) visited

b) have visited

c) were visited

d) visit

(29) The wings are _____ to the fuselage.

a) attaching

b) attached

c) attaches

d) attach

(30) _____ it was necessary to attend the lecture on Aeroengine, neither Ali nor Abdulazaz did.

a) However

b) But

c) Although

d) Nevertheless

(31) Is that the runway _____ the aircraft stopped?

a) which

b) where

c) when

d) were

(32) He _____ swimming regularly, but now he never goes.

a) was used to go

b) used to go

c) use to go

d) goes

(33) He was the best pilot _____ them all.

a) of

b) from

c) by

d) than

(34) The pilot _____ life was saved is my best friend.

a) whose

b) of whom

c) when

d) of which

(35) He left for Jeddah _____ Tuesday.

a) at

b) on

c) by

d) in

(36) I have read the book _____ you lent me.

a) that

b) whom

c) whose

d) what

(37) If he had flown at a lower level he _____ avoided the birdstrike.

a) could had

b) would have

c) will have

d) should have

(38) When the wing stalls there is a _____ in lift and a increase in drag.

a) reduction

b) reduced

c) reduce

d) reducing

(39) As the pilot _____ the runway, he saw an aeroplane taking off.

a) approach

b) was approaching

c) is approaching

d) approaching

(40) "Have you flown solo?" I asked my roommate.

I asked my roommate _____

a) had he flown solo

b) if he had flown solo

c) that he had flown solo

d) had you flown solo

(41) Did you meet the pilot _____ wife gave birth to a baby boy.

a) whose

b) that

c) whom

d) which

(42) If the air hadn't been turbulent the pilot _____ to climb.

a) wouldn't had decided

b) wouldn't decided

c) wouldn't have decided

d) wouldn't has decided

(43) The temperature of air usually decreases _____ altitude.

a) in

b) by

c) into

d) with

(44) "By the end of next month _____ solo".

a) we will be flying

b) we would fly

c) we have flown

d) we fly

(45) "Where is your car?"

"It _____ in the garage"

a) is being repaired

b) is repaired

c) is repairing

d) is been repaired

(46) The Strikemaster is longer than the Cessna _____ five feet.

a) by

b) in

c) with

d) of

(47) This is _____ interesting book I have ever read.

a) most

b) the most

c) a most

d) the more

(48) "Are there many students in the lecture hall?"

"No, only _____ turned up."

a) a few

b) less

c) little

d) a little

(49) _____ you are sick you must stay in bed.

a) Although

b) Therefore

c) However

d) Because

(50) Yesterday we watched a film _____ flying.

a) about

b) around

c) of

d) on

(51) We _____ of the examination results two hours ago.

a) had informed

b) have informed

c) are informed

d) were informed

(52) Air traffic control officers give aircraft regular up-dates of weather condiditions _____ the airport.

a) in

b) on

c) at

d) about

(53) You _____ get your hair cut before the officer sees you.

a) should rather

b) could

c) should

d) may

(54) Water boils _____ you heat it to 100C°.

a) unless

b) if

c) until

d) although

(55) If he had worked hard he _____ passsed the examination.

a) would have

b) could have

c) will have

d) had had

(56) "He speaks very good English"

"Yes he does. He _____ in England for three years.

a) was learning the language

b) has been learning the language

c) has learned the language

d) have been learning the language

(57) Choose the correct sentence.

a) A first plane I flew was Cessna 172.

b) The first plane I flew was a Cessna 172.

c) A first plane I flew was a Cessna 172.

d) The first plane I flew was Cessna 172.

(58) "Is this your log book?"

"Yes, this logbook is _____"

a) mine

b) yours

c) my

d) your

(59) "Where have you been? We have been waiting for you

_____ 12 o'clock"

a) for

b) during

c) between

d) since

(60) This palmtree _____ in 1982.

a) planted

b) plants

c) is planted

d) was planted

(61) The instructor took us to see the newly _____ plane

a) manufacture

b) manufacturing

c) manufactured

d) manufacturer

(62) Would you be interested in _____ a lecture on Mathematics?

a) attended

b) attending

c) to attend

d) attend

Test 2

Reading Comprehension Test

This is a test of your understanding of written English. Here are four passages. In each passage a number of words are deleted. The deleted words are replaced with short lines _____. Read the whole passage first then go back and fill in the blanks with the one word you think is missing.

NAME: DATE:

SERVICE NUMBER:

COURSE:

A.

Instructions:
Read the whole passage first. Then go back and fill in the blanks with the one word you think is missing. The following is an example:

I was born on Tuesday the 25th of June

An aircraft moving on the ground by its own power is said to be taxying. Taxiways are specially prepared paths _____ the airfield for use by _____ aircraft. They are made of _____ same materials as the runways, _____ must be of the same _____. They connect the runways to _____ dispersals and aprons. The centre _____ the taxiway is marked by _____ broken yellow line. This is _____ as a guide to pilots _____ taxying aircraft.

At the place _____ a taxiway joins a runway, _____ continuous white line is painted _____ the taxiway. This line, known _____ the holding point, this is _____ an aircraft must wait until it has received permission to enter the runway.

TEST OF ENGLISH FOR AIR CADETS

T.E.A.C.

TEST 2

NAME: DATE:

SERVICE NUMBER:

COURSE:

B.

Instructions

Read the whole passage first. Then go back and fill in the blanks with the one word you think is missing. The following is an example.

I was born on Tuesday the 25th of June.

Fog is a very common example of poor visibility but here in Saudi Arabia is rarely seen. In Riyadh it may _____ occur about once or _____ each winter, mainly because _____ air is too dry _____. When fog does occur _____ pilot should be able _____ recognize it and understand _____ it forms.

Fog is_____ simply a cloud at _____ level. It is made _____ many tiny water droplets _____ are suspended in the _____ reducing the horizontal visibility _____ less than 100 metres, _____. When there is very _____ stratus cloud, at about 100 feet, and the wind becomes very light, the stratus cloud lowers onto the ground to make fog.

NAME: DATE:

SERVICE NUMBER:

COURSE:

C.

Instructions

Read the whole passage first. Then go back and fill in the blanks with the one word you think is missing. The following is an example.

I was born on Tuesday the 25th of June.

Air Traffic Controllers are men who keep in contact with aircraft in the air.

Some of them work _____ the control tower from _____ they can see everything _____ happens on the airfield. _____ a pilot wants to _____, Air Traffic Control tells _____ which runway to use. _____ work in large control _____ from which they monitor _____ position of aircraft by _____. If contact with an _____ is lost Air Traffic _____ will ask other controllers _____ help locate it.

All _____ Traffic Control Centres are _____ by direct telephone and teleprinter links. Sometimes they communicate by radio. Contact with aircraft is by radio.

NAME: DATE:

SERVICE NUMBER:

COURSE:

D.

Instructions

Read the whole passage first. Then go back and fill in the blanks with the one word you think is missing. The following is an example.

I was born on Tuesday the 25th of June.

In old aircraft the pilot's position was not covered. We say that they had open cockpits. When the pilot climbs _____ his position from above _____ call it a cockpit. _____ an open cockpit the _____ must wear goggles and _____ helmet to protect him _____ the airstream.

Today's aircraft _____ covered cockpits. The pilots _____ wear goggles. They wear _____ oxygen mask and a _____ helmet. They need oxygen _____ they fly very high. _____ the oxygen mask there _____ a built-in microphone so _____ the pilot can speak _____ his radio. In his helmet there are ear pieces so he can listen to messages from his radio.

Test Three

Listening Comprehension

Part One

This is a dictation. You will be given two paragraphs to write down. The first time you hear each paragraph, just listen. The second time there will be pauses for you to write down what you hear. The third time you will be given marks of punctuation. When you hear 'full stop' do not write down f-u-l-l - s-t-o-p but put down a dot.

It is important*6 that a pilot 4 should know something 8 about the many components 10 of his aircraft. 4 At first 3 your Instructor 5 will point out 6 only the main features 12 but, as the course progresses, 12 you will be able to discuss 13 your aircraft 5 and its components 7 in more detail.6 It is important for you 10 to do this 3 with confidence, 3 particularly when you have to report 15 aircraft unserviceability 9 to the servicing personnel.10

Your Instructor 5 is well qualified 7 to help you 3 become a competent Service pilot.19 You can learn 5 by his example 5 as well as 3 from his instruction, 6 so watch carefully 7 when he is controlling the aircraft.16 If you do not understand 12 why your Instructor 10 does something 4 while flying the aircraft,9 ask him.2 Never be afraid to ask questions;11 we all have to learn.8 By asking questions 6 you can learn 5 from the experience of your Instructor.15

* numbers denote length of pauses in seconds

Listening Comprehension

Part Two

In this test you will hear a lecture. You will hear it only once. You may make notes on this side of the answer paper as you listen. You must not look at the questions until the end of the lecture. The questions are on the other side of the paper.

Notes

Good afternoon.
Today I'm going to talk about the subject of Meteorology. I shall introduce you to the idea of what Meteorology is and why it is important for you, the pilot, to have a sound knowledge and understanding of the subject.

Meteorology is one the earths' sciences. It is a science about the earths' atmosphere and a study of the clouds and the weather. The earths' sciences also include Geography which is the science about the diverse features of the earth's surface, its peoples and the crops they grow; Geology which deals with the history and composition of the earth and rocks; Oceanography, a study of the oceans and the waves and Climatology, the science about the many different regional climates of the world and its contributions to the environment of life.

Weather itself has different meanings but basically it is used to describe what you sometimes see falling from the sky like rain, snow and hail, how far you can see along the ground or just simply how warm or cold you feel. One thing you will almost certainly have noticed is that the weather changes from time to time, sometimes very slowly but on other occasions it changes very quickly, in a matter of a few minutes. These weather changes happen because of changes which are going on in the atmosphere and because these changes are taking place all the times it is necessary to follow the weather not just at the present time but also to have an idea about how the weather will behave in the future. Weather in fact is one of the many aspects of life from which it is difficult to escape. The weather affects all of us and whether we are at school, university, at home or at work, all of us take some interest in the weather. Fishermen for example dare not go to sea when the winds are strong and the seas rough, mountain climbers look for shelter when the clouds are low and it is very cold while farmers look for warm and sunny weather when it is time to gather the harvest. However few are more affected than are pilots who have to face the hazards of the weather on the ground and in the air whether they are involved in military, civil or transport flying. Even with todays modern jet aircraft and its sophisticated instruments, aircraft may still be delayed, re-routed or even damaged by the sometimes very unpredictable changes which occur in the atmosphere. For example pilots need to know about the likelyhood and time of onset of severe weather conditions such as thunderstorm, icing, turbulence, squally winds and tropical revolving storms so that they can plan their routes and sorties effectively with at the same time minimum discomfort to the crew and passengers. Many of us have flown at one time or another and know only too well the tension and distress we can feel when the aircraft is being buffetted about by turbulence or when you are in thick cloud and cannot see the ground. Whichever form of transport we use, we always look and hope for a smooth journey and the captain of the aircraft with his expertise and knowledge about the weather does the same, to try and ensure a safe and comfortable flight.

Fuel economy is also a very important consideration. When you drive your car on a long journey, you often look for the shortest route or a way which avoids delays and consumption of fuel. Likewise, a pilot on a long haul flight will plan his route according to the winds. Many times we hear about Westbound flights over Southern Asia and the Atlantic Ocean taking one to two hours longer than a corresponding flight in the opposite direction. This is caused by the presence of very strong winds at cruise levels, usually between 30,000 feet and 40,000 feet in the atmosphere. These very strong Westerly winds which are often blowing at more than 100 knots are called a Jet Stream and as well as fuel considerations they create turbulence and generate weather. For these reasons then and many more, aircrew and operational staff should have a sound and thorough knowledge of the principles of Meteorology, the clouds and the weather so that they can follow and understand all the information given to them at aircrew and meteorological briefings.

At this stage you may be asking yourselves what kinds of weather information are available and where does the information come from. Well! it's a long story but in recent times and especially during the last two decades, enormous progress has been made with the reliability and accuracy of weather reports and the methods of transmission of weather data. It may be useful at this point to say that details about the present weather conditions are often referred to as a weather report or meteorological situation while an interpretation of what weather is expected to happen in the future is called a weather forecast. Many years ago before sophisticated instruments were available to measure and record such elements of the weather like temperature, pressure and wind people often kept written records or diaries of their local weather. These people were often landowners or landed gentry who had plenty of time to marvel at the skies and the ever changing moods of the weather. This I suppose really was the origin of the amateur weatherman who at times claims that his way of monitoring and forecasting the weather is just as good as the methods used by highly trained professional personnel and their expensive machines. Even today many enthusiastic amateur weathermen keep records of the weather, and their detailed accounts of it are often very useful in weather analysis. Some people have learned how to read and foretell the weather in other ways. A rural farmer for example will often be able to tell you when he thinks it is going to rain or alternatively the prospects for fine weather by looking at the sky simply because he spends most of his time out of doors and becomes very familiar with the appearance of the sky and the accompanying weather. Many hundreds of years ago and right up to the present day in fact, some folk use old fashioned rhymes which we call weatherlore. One of these for example is:

Red sky at night, shepherds delight
Red sky in the morning, shepherds warning

This is supposed to mean that if you see a red sky at sunrise or soon afterwards, it will rain that same day whereas a red sunset is

supposed to foretell a fine day to follow.

There is no scientific knowledge to support this idea and although it is believed to be accurate for about 70% of the time it can only be applied to Temperate climates where there is sufficient rainfall to make it a plausible statement.

Whichever it be, weatherlore or a subjective appraisal of the weather by the amateur weatherman or landed gentry alike, it is of little use when it comes to planning a flight or deciding when to set sail. In some European countries, gas and electricity corporations receive temperature forecasts so that they can assess how much fuel is likely to be consumed within a given time while at sea, ships are routed to avoid the storms and rough waters in order to preserve their cargo. All of this would not be possible without a large number of meteorological personnel who measure physical data and use computers which make millions of calculations using complex mathematical equations every second. These computers produce forecasts very quickly and are much more detailed and reliable than their weatherlore predecessors but they rely on accurate information being fed in to them. The familiar computer saying "garbage in, garbage out" is very true, because you can only expect to receive a correct and useful forecast if you give it the right and accurate information in the first place.

To do this many observations of the weather have to be made at regular intervals, on land, at sea, in the air and from space. On land a large network of meteorological stations observe and record such elements as weather, temperature, pressure, visibility and wind and observations of the amount and height of each different type of cloud are also made. In addition there are automatic weather stations and many voluntary observers to assist the network of observing stations. Ships at sea, including the merchant fleet, also send in their reports while in the air aircraft reports of cloud formations, turbulence and thunderstorm can be extremely useful. In space satellites, some orbitting the Poles and others almost stationary high above the Equator, send back to earth pictures of the clouds from above. These pictures reveal the masses of cloud associated with bands of weather and by comparing the difference in time between two orbits it is possible to get a good idea of how fast the cloud and weather is moving. Additionally the infra red detectors help us to distinguish the white and higher clouds at lower temperatures from the darker, warmer and lower clouds which are closer to the ground.

Observations of the weather are usually made on the hour and in many cases every half hour at busy international airports and the next step is to collect and redistribute all the observations and weather data. To do this observations are first collected by meteorological centres within a country and then they are sent to International collecting centres such as Frankfurt, Washington, Paris and Bracknell. In other words information about the weather is collected on a world wide basis; fortunately weather has no

political boundaries and data is freely interchanged between one country and another.

For this purpose an organisation has been set up to produce standard meteorological codes and reporting procedures so that everyone no matter what language they speak can understand the weather data. It is called the World Meteorological Organisation and it is based in Geneva in Switzerland. As a part of this set up each country has a central forecasting office through which data is collected and exchanged and in addition other stations may be responsible for military and civil aviation. Transmission of the data is also a very important part of the chain because if the weather information is going to be used then it must be available readily and quickly. Many years ago radio was the only means of communication but these days we are able to rely on telex, computers and satellites to transmit the data around the world in a matter of minutes. This enables detailed and more accurate forecasts to be made and as a routine forecasts are usually updated frequently to take care of any unexpected changes in the weather.

Weather observations and the collection and dissemination of data are carried out day and night, every week of the year for it is a continuous occupation. It is a most complex business, gentlemen, but it is a challenge to mankind and a very interesting subject to follow. This then is the organisation of the meteorological operations and forecasting. It is very true to say that Meteorology has progressed a long way since the days of weatherlore and feather pens and since the arrival of the computer age during the last 20 years great strides have been made to provide up to date weather data and forecasts with a very high standard of accuracy. Today it is almost an exact science with only a little bit left for chance. Fishermen, sailors, mountain walkers, farmers and pilots and many other people in different walks of life can now place a fair amount of confidence in the daily weather forecast.

In the next talk, I'll look at Meteorology purely as a specific science and we'll study what causes the weather and why it changes from day to day.

Good-day gentlemen.

This is the end of the lecture.
Now turn over to the questions. Answer them quickly.

(1) Meteorology is one of the earths' sciences and is a study about:

a) the earth's surface

b) the oceans and the waves

c) the many regional and different climates

d) the clouds and the weather

(2) Meteorology affects:

a) only a few people

b) almost everyone of us

c) only military and civil pilots

d) only farmers, fishermen and pilots

(3) An aircraft flying across the Atlantic Ocean with a head wind of more than 100 knots is likely to:

a) arrive on time

b) arrive ahead of schedule

c) be about 2 hours late on arrival

d) be re-routed but still arrive on time

(4) A Jet Stream is:

a) an Easterly wind

b) a squally wind

c) a wind blowing at 100 knots or more

d) the wind at cruise level

(5) A red sunrise in the morning means that:

a) there's about a 70% chance of rain during that day

b) a red sunset will follow later in the day

c) it will soon be raining

d) the next day will be fine

(6) The World Meteorological Organisation is based at:

a) Frankfurt

b) Washington

c) Paris

d) Geneva

(7) A computer used for weather forecasting:

a) will always produce a correct weather forecast

b) relies on accurate information to be fed into it before the forecast is made

c) takes a long time to produce a forecast

d) relies on weatherlore for some of its information

(8) A satellite provides weather information about:

a) the wind

b) turbulence

c) the clouds

d) the visibility

(9) The weather changes:

a) at regular intervals of time

b) always very slowly

c) only during the daytime

d) sometimes very quickly

T.E.A.C.

TEST OF SPEAKING

An Interview

Content and stages

Stage I: Introductory

a. Greeting the interviewee

b. A statement by the interviewer explaining the purpose of the interview

c. Preparation for the following stage making sure that the interviewee is sitting at ease, and all other administrative work and personal details of the interviewee is checked and completed.

Stage II: Personal background (in the beginning mainly wh questions)

a. Home province

b. Education

c. Father's occupation

d. Travel abroad, purpose

e. Reasons for joining the Academy

f. Hobbies

Stage III: a. A description of a typical day in the Academy

b. A comparison between life in the barracks and life at home

c. A description of a typical weekend

d. A description of the interviewee's home town

e. A description of an aircraft

f. A description of an airport

Stage IV: * - Asking a question or questions by the interviewee

Stage V: - Closing the interview

- Thanking the interviewee

* The purpose is to assess the testee's ability to form questions.

TEST OF SPEAKING

ASSESSMENT SCALE

Cadet's name: Date:

Service number: Course:

Level 7. Native-speaker ability.

Level 6. Near native-speaker ability and fluency with occasional evidence of non-native speaker pronunciation features.

Level 5. Full communication established, but with occasional minor faults of English usage and pronunciation.

Level 4. Adequate communication, sometimes impaired by a number of minor faults of English usage with occasional evidence of hesitations.

Level 3. Communication exists but dominated by hesitations and frequent errors in grammar/lexical choice and pronunciation.

Level 2. Communication is limited to answering 'wh' questions with serious errors in grammar and pronunciation.

Level 1. No communication at all.

ACADEMIC SKILLS QUESTIONNAIRE

Instructions: Please number choices from 1-----4, with 1 as important

A. Of the major skill categories, which are most essential to success in your classes? PLEASE CHECK ONE OR MORE

_____ Reading (e.g. of texts)
_____ Writing (e.g. of reports, papers)
_____ Speaking (e.g. oral reports, discussions)
_____ Listening (e.g. to lectures)

B. Reading: which reading ACTIVITIES are ESSENTIAL? Reading of:

_____ Textbooks
_____ Examination questions (multiple choice)
_____ Examination questions (essay)
_____ Non-textbook (simplified texts and quiz questions prepared by instructor) assignments

other _____

- -

C. Writing: which writing ACTIVITIES are ESSENTIAL? writing of

_____ REPORTS
_____ TERM PAPERS
_____ ESSAY ANSWERS

OTHER _____

- -

D. Speaking: which speaking ACTIVITIES are most essential?

_____ Ability to participate in class discussion
_____ Ability to formulate questions
_____ Ability to organise and present an oral report
_____ Ability to pronounce words clearly

other _____

- -

E. Listening: which listening ACTIVITIES are ESSENTIAL?

_____ Ability to follow spoken directions
_____ Ability to comprehend and restate orally-presented classroom material
_____ Ability to understand and take adequate notes on lectures

other _____

WHICH IS MORE IMPORTANT FOR SUCCESS IN YOUR CLASSES, A PROFICIENCY IN GENERAL ENGLISH OR A PROFICIENCY SPECIFIC TO DISCIPLINE? PLEASE SPECIFY

Your subject _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

comments _____

For further comments please turn page over.

ENGLISH ABILITY RATING

Key for completing the accompanying English ability rating form.

(1) Please complete one form for each student.

(2) The information requested should be supplied in the boxes on the accompanying rating form.

Please note that the Rating forms assess the present ability.

Thank you for your co-operation.

ENGLISH ABILITY RATING

Cadet's name: Date:

Course:

1. General ability in English (put a in the appropriate box in column A and B)

points	A		B	
6	☐	Excellent i.e.	☐	Shows near-native speaker ability. Has no difficulty in expression or understanding
5	☐	Very good i.e.	☐	Clearly a non-native speaker because of minor faults in English usage, but this does not handicap him in his studies. Makes complete communication
4	☐	Good i.e.	☐	Makes many mistakes in English, but this constitutes only a minor handicap for him in his studies.
3	☐	Satisfactory i.e.	☐	Shows many weaknesses in English but his English ability can be considered just adequate for his studies. A higher standard is desirable.
2	☐	Weak i.e.	☐	Shows very little ability in English both in expression and understanding and is well below a satisfactory standard.
1	☐	Very weak i.e.	☐	Shows no ability, no communication.

2. Individual language skills

	Excel-lent	V.good	Good	Satis-factory	Weak	Very Weak
Ability to understand spoken English						
Ability to speak English						
Ability to understand written English						

Any further comments: please specify

(A)

Deleted Items in the cloze

Source: Airmanship students' lecture notes
King Faisal Air Academy

(1)	On	(6)	the	(11)	where
(2)	taxiying	(7)	of	(12)	a
(3)	the	(8)	a	(13)	across
(4)	and	(9)	used	(14)	as
(5)	strength	(10)	of	(15)	where

(B)

Source: Meteorology students' lecture notes

(1)	only	(6)	to	(11)	which
(2)	twice	(7)	how	(12)	air
(3)	the	(8)	very	(13)	to
(4)	Nevertheless	(9)	ground	(14)	sometimes
(5)	the	(10)	of	(15)	low

(C)

General Aviation English, School of English Air Service Training

(1)	in	(6)	him	(11)	aircraft
(2)	where	(7)	others	(12)	Control
(3)	that	(8)	centres	(13)	to
(4)	When	(9)	the	(14)	Air
(5)	land	(10)	radar	(15)	connected

(D)

Source: General Aviation English Air Service Training

(1)	into	(6)	from	(11)	because
(2)	we	(7)	have	(12)	In
(3)	In	(8)	don't	(13)	is
(4)	pilot	(9)	an	(14)	that
(5)	a	(10)	flying	(15)	into

1. Observation and experience of the process of teaching technical subject-matter in the students' native language suggests that problems of comprehension and expression that are encountered when the medium of instruction is a foreign language are due mainly to lack of proficiency in understanding and using the basic vocabulary and structures of the foreign language, and not to any difficulty in the subject-matter itself. I would therefore expect all of the Aeroscience instructors who respond to this questionnaire to say that proficiency in general English is more important for success in their classes than familiarity with the concepts and technical language of the subjects they teach.

2. Insofar as learning problems in the Academic Wing are linguistic, they arise from lack of familiarity with the basic vocabulary and structures of non-technical English and, of these two elements, lack of familiarity with the basic language structures is the more important.

3. The English Department can contribute to an improvememt in the quality of Aeroscience and flying training mainly by making basic ELT more efficient, and not by diverting more time and resources than at present to the teaching of the technical languages of the post-ELT phases of training.

4. How then can the efficiency of our basic ELT be improved? Obviously, all aspects of any training programme can always be improved to some extent. The biggest improvements, however, can be made by aiming to eliminate the bigget and most obvious weaknesses. In my opinion, the main weaknesses of the present ELT programme are these:
(1) ELT is pursued within an overall training framework which has the effect of sending the cadets to ELT classes when they are too tired to profit much from the instruction.
(2) The method of testing in ELT, which determines the directin in which both students and instructors will expend their main effort does not motivate the students to acquire useful language skills.

5. Both of these problems should be faced, but for now let's consider only the second. If you want the cadets to be able to understand easily explanations in basic English, to respond quickly and appropriately to instructions, to speak fluently, to formulate questions easily, and so on, you must make it worth their while to acquire these skills by testing each such skill and by insisting that the desired level of proficiency in each area is achieved before the students leave full-time ELT. Therefore I would say that the first thing to be done is to redesign the testing system so that the whole ELT programme becomes more specifically objective-oriented.

Comments by Derek Nichols

PHONEMIC TRANSCRIPTION OF SOME OF THE ORAL INTERVIEWS

Oral Interview - this interview denotes level 5 Content

* dots denote pausing

Interviewer: Right Nasser, if you like to sit down. Now, did your family come from Riyadh?

Cadet: Oh yes they came from Riyadh.

I: Has your family always lived in Riyadh?

C: No, first of all they came from Ạlgaseen then they stayed at Riyadh since nineteen ..... nineteen sixty three.

I: So you have always lived in Riyadh.

C: Yah

I: Which part of Riyadh do you live in - Nasseryah or Sulaimania?

C: In Alolyal.

I: Olya, I see. Your family have a big house or a flat?

C: No, they have a palace.

I: A palace! Good heavens!!

C: Near the centre of Olya.

I: Right, O.K. So you went to school in Riyadh?

C: Yah

I: Can you tell me something about the schools that you went to in Riyadh; how old were you when you started the primary school?

C: I started the primary school when I was six years then I spent six years - of course - in the primary school. After then I went ..... (Cadet was interrrupted by the Interviewer.)

I: Secondary school.

C: (Continuing) .... secondary school for three years more. After than to the high school.

I: What subjects did you study at high school?

C: At the high school - of course I am a scientific you know - there I had to study the mathematics, the physics and many things like that.

I: Chemistry?

C: Yah.

I: Biology?

C: No.

I: Mathematics and physics?

C: Yah.

I: And what about English. When did you start learning English?

C: I started learning the English language at the secondary school.

I: When you were what? 12 - 13?

C: About 12 years old.

I: Your English teachers, were they from England or were they from Egypt?

C: First of all I had a teacher who was from Saudi Arabia then at the second year of the secondary school I had a Sudanese teacher.

I: You had, what, one English class every day, five hours a week?

C: No. I have ..... Sometimes we have two lessons every day you know.

I: Right! Good! Well! Your father's occupation.

C: Yah.

I: You live in a palace in Riyadh.

C: Yah.

I: What does, did, your father do? He must be important.

C: No, he is a business man. You know he is doing a business all the time.

I: What kind of business is he doing? Import export?

C: All these things importing ..... import and export and sale the land and something like that.

I: How do you come to live in a palace? Is this -

C: It's exciting really.

I: Yah, but I mean, can anybody live in a palace or do you have to be a member of - connected to the Royal family? Or to be somebody important?

C: No, it depends on the money. If you have enough money to or build a palace you can build it and stay in it.

I: Can you describe this palace in Olya because I would be interested to see it.

C: O.K. it's about six thousand meters ..... square meters, and there is in the middle of it, this building, I mean the palace building, and then about 9 room and then is ..... (interrupted)

I: Nine rooms in the palace! Quite small palace?

C: No, these nine rooms, its only for sleeping.

I: Nine bedrooms?

C: Yah nine bedrooms.
..... and there is six bathrooms and two kitchens and .....

I: How many floors?

C: Oh it's only two floors.

I: Do you have any help, any servants, in the palace?

C: Yah.

I: Six bathrooms and nine bedrooms ..... there is a lot of work to do.

C: Yah I know that.
Then I have one. She is from Thailand and she is a helper of my mum you know and we have of course the driver and the cleaner of the house.

I: Right. Now visits abroad. Have you ever been abroad? Have you been outside Saudi Arabia?

C: Yah, I have been to many countries.

I: Can you tell me some of the countries you have been to?

C: I have been to your country, I mean to England ..... if you know Exeter.

I: In Devon.

C: Yah in Devon. It's the head of Devon I think.

I: It's the capital.

C: The capital.

I: Yes that's right.

C: You know and I have been studying the English language over there, in "Siarah Academy" and the International School, if you heard about it. Have you been there?

I: There are international schools in different places in England. I have not been to international school in Devon but I have been to international school in Piccadilly.

C: Yah.

I: International House School.
How long did you study English in Devon?

C: About three months.

I: Three months. Did you live with an English family when you went there?

C: Yah.

I: And you studied English in a class, what, international class, men and women?

C: Yah. Of course and there were six people - Japanese - from France and two people from Japan and one from Saudi Arabia, and two from Spain.

I: Did you find it difficult to get used to living in England after Saudi Arabia?

C: Of course, first of all there is a difference in the customs and something like that and I have to get myself used to live with the English people you know and to the different food of course because all that time, they only gave me french fries I mean chips and hamburger and something like that.

I: Right, right. What about things like bacon. You did not have bacon for breakfast?

C: Yes of course, they gave me but I said to them no please because our religion says that if you would like to eat in the morning you have to eat something good like cheese or something like that you know.

I: Good. What other countries have you been to?

C: To America.

I: You have been to America, with your father or did you go alone?

C: No on myself.

I: Just by yourself.

C: Yah.

I: And what did you travel by? Greyhound bus or did you fly to different cities?

C: No, first of all I would like to speak about my trip to America. I went from Riyadh to Jeddah, then from Jeddah to New York, I arrived there at Lagardia airport then I changed this airport to other airport I think its Kennedy airport then I took the Pan Am from New York to Los Angeles then I stayed there and I have been there three months over there.

I: In a hotel or in a flat or what?

C: No in a flat.

I: By yourself?

C: Yah, but I have my friend over there. He is my cousin you know, and he is studying over there.

I: Good, and you have been to other European countries or is it mainly the Middle East like Egypt and Jordan?

C: Yah, Egypt and something like that.

I: Right. Good. Well now obviously you have come to the Academy because you want to become a pilot. Have you always wanted to be a pilot?

C: Not at all. When I saw some bad accident or something like that I am afraid - of - to - be a pilot but when I saw the moving of the aircraft like the F15 moving and doing the loop to the left and to the right I find it fun flying the F15 or something like that.

I: Did you come here to the Academy because you wanted to come or did your father recommend it?

C: No, I came here because I wanted to come here and my father said if you would like to be here or at the University it's up to you. He said that you know then I decided it's better if I come here and be a pilot.

I: How long will it be now before you take a go out in the Cessna? Start your flying?

C: I think six months.

I: Another six months.
You passed the E.C.L. and now you are doing Aeroscience?

C: Yah, that's right.

I: Would you like perhaps to describe a typical day in the Academy?

C: O.K. In the morning we wake up at three and half.

I: Half past three.

C: Yah, after then we have about fifteen minutes washing our faces and cleaning our rooms you know, then we have to go to the mosque to pray then we have, it's about four o'clock 'the prayer time'. After then we have to go to the tarmac to do some exercises, you know, then we will go to the 'Mase' I mean the restaurant over there to have the breakfast and after then we have to go to the barracks again to take the books and at six o'clock we have to go to the Academy to study, we, then, finish about twelve o'clock after then we will go back again to the barracks having our lunch and so on then we will go back to the barracks to stay and have some sleep, after then, if one would like to study he will have a private study over there and after then about six o'clock we will have something called (I don't know it in English) but I think it's like to know how many people are in the Academy right now or is there anybody missing.

I: Oh yes. What do you do in the afternoon? I mean between four o'clock and six o'clock. Do you play sports yourself?

C: Yah you mean after four o'clock we have two things to do and it keeps changing all the days first of all we have an army exercise and the other day we'll have a

sport exercise.

I: And then in the evening after supper no time is there?

C: After seven o'clock we'll have our dinner after that we'll go back to the barracks studying and something like that, to clean the rooms and be careful that everything is all right in the barracks. We then go to sleep about nine o'clock.

I: How long were you learning English in the Academy before you took the E.C.L.?

C: One month and a half.

I: Only one a half months.

C: Yah.

I: Because you have learnt English in England and America and you used the A.L.C.?

C: Yes of course.

I: Was it good?

C: Oh really - it's good from one side and bad from the other side. I mean it's bad because you will not learn the conversation and you will not learn spelling and it's good because you will finish this course very fast.

I: So when you were learning English in Exeter in the language school in Exeter, was this better? The system there was better than it is in the Academy here?

C: Really in Exeter it's much better because of one thing you know. Because you were there speaking in English all the time only with the English people and the French people then you will have to learn the language without knowing that, because when you were at school you have to speak the English language. In the street you have to speak the English language and at the family you will find the English people over there, you will speak in Arabic and the T.V. It is all about the English language then you learn more.

I: Of course. If you go home here to the barracks you will speak Arabic.
What about weekends. When you have a long weekend from the barracks here. Do you go to Riyadh, to your palace in Olya?

C: Yah.

I: How do you spend a typical weekend? What kinds of things do you do?

C: It depends. If I have a test or something like that or a final exam, of course I will be studying all the time. If I haven't anything to do I'll go with my family on vacation or something like that.

I: Do you have any hobbies or interest? For example you go bowling in the bowling club or you go skating at the supermarket in Olya.

C: We have a swimming pool at our palace you know and during that time I swim over there and play tennis and something like that.

I: What about Thursday night though? I always wonder what people do on Thursday. Thursday night is Saturday night in England and America. What do you do on Thursday night?

C: Maybe I'll go with my friends to the desert having some games and of course we will do some sport. We will also eat some food. We put something on the fire, like a steak.

I: Barbecue.

C: Oh yes.

I: Do you like the life in America? You were there for three months in Los Angeles. Do you like it there or do you prefer the life in Saudi Arabia?

C. Really I prefer the life in Saudi Arabia.

I: O.K. Are there just one or two questions you would like to ask me at the end. Any questions.

C: Before I would like to ask you about your name.

I: My name is Robert. My given name is Robert and family name is Triggs. Robert Triggs.

C: What does that mean? I mean Triggs of course is different about the Arabic language and English language, because Triggs in Arabic is strange name, but if you have any meanings for that.

I: Well, it only means that the first three letters

T-R-Icome from Cornwall near Devon. People who come from Cornwall, many people have names which begin with Pen. So you get lots of names beginning with Tri and Pen. My family came from Cornwall 100 years ago so that's why I am called Tri.

C: And, were do you come from then?

I: Well I come from London.

C: From London itself.

I: Yes from London, middle of London. I was born in London but my family one hundred years ago lived in Cornwall next to Devon.

C: And how do you like it. I mean how do you like to live in London with all this crowd and .....

I: Well, I like to get out of London as much as I can at weekends.
O.K. thank's very much, that is very good Nasser.
Thank you.

Phonemic transcription of part of an oral interview

denoting Level 3

Interviewer: All right Hamond, take a seat please. Ah, do your family come from Riyadh?

Cadet: No ..... they are live in South the Kingdom.

I: South Abha or .....

C: Near Abha. Village near Abha. It is called (ALMITALI).

I: How far is that from Abha?

C: From Abha about 100 Km.

I: Has your family always lived there?

C: Yes yes.

I: Did you go to school in Abha or in your village?

C: No - no in my village.

I: What! How many people live in your vIllage? What's the population?

C: About a thousand.

I: Only a thousand.

C: Yes.

I: And you've got a primary school, a secondary school - high school?

C: Yes yes all of them in my village.

I: Can you just tell us something about your education. How old were you when you went to school?

C: Yes yes. I begin my studying about 14 years ago. I begin the primary school of course. I did not return on it and begin secondary school and after that high school.

I: What did you study at high school?

C: I didn't take scientific subjects.

I: Sorry, you didn't take or you did take?

C: I did not take.

I: So what did you take then?

C: Some subjects about geography and history and English of course. English was four lesson in the week.

I: Four lessons a week.

C: Four lessons a week.

I: What does your father do?

C: He is a farmer.

I: A farmer. Does he have a big farm?

C: No no he small.

I: So what does he have on the farm? Animals, or is it mainly crops?

C: No crops and a little animals, few animals.

I: For example.

C: For example cows just cows.

I: And do you help him on the farm? Have you helped him?

C: Yes of course.

I: Would you like to do this kind of work yourself?

C: Yes our job. First our job.

I: But you've decided to become a pilot. Because of what?

C: Yes. Because it is first to become pilot.

I: Have you been outside Saudi Arabia?

C: No I didn't.

I: Never.

C: Never.

I: Would you like to go to America or England?

C: No I don't like.

I: If the Academy sent you and said we would like you Hamond to go to Texas would you like to go or not?

C: No I wouldn't like to go. I would like to stay here.

I: How long have you wanted to be a pilot?

C: From ..... from ..... when I in primary school.

I: Why was that, becuse you saw a film that interested you or what?

C: No no I do not but its first.

I: You have not got brothers who are at the Academy?

C: No no.

I: Perhaps you would like to describe a typical day in the Academy as Nasser did.

C: We get up about half past three and we prepared for praying about 15 minutes to wash our face and prepare our rooms. After that we go to a mosque to praying, when we have finished from praying we return to make the royal call after that we have running about the tarmacs after that we go to mase. When we have finished from the breakfast we return to the barracks to take our books. After that we come to the Academic Wing. When we have finished exactly 12 o'clock we return to the barracks to prepare for the praying. We go to the mosque about half past twelve. When we have finished about one o'clock we go to the cadets' mase. When we have lunch there, when we have finished we return to the barracks. If there is any cadet have punishment he go to the tarmac for one hours. After about three o'clock we go to the mosque for praying then we return to the sport or the army practice. Some squadron have sport and the other have army practice. We go about four o'clock, when we have about six o'clock we return to the barracks and prepare to the royal call. After that we go to the cadets mase to take our dinner when we have finished we return to the barracks and prepare to the fifth praying. We return to the barracks for sleeping at nine o'clock.

I: Nine o'clock ..... so when you first came to the barracks from your village in the south of Saudi Arabia, was it very difficult to get used to the life here?

C: No I didn't get any difficulty.

I: But the life here is very different from your life in your village. What's the difference? One or two differences.

C: The difference first of all getting up early and ..... staying in the barracks for a long time ..... we didn't go outside the academy, we can do everything with ourself.

I: Right. What are your hobbies, do you have any particular hobbies?

C: Yes, reading the old books and playing football.

I: You said reading old muslim books.

C: Yes old muslim books.

I: O.K. Well now, what about your typical weekend. When you go back to your village .....

C: No, no, we can't go there. I go to my brother in Riyadh.

I: He has a flat in Riyadh.

C: Yes yes.

I: What does your brother do? Is he a business man?

C: No he is a major in the Army.

I: In the army, I see. O.K. Are there any questions you'd like to ask me?

C: Yes. What's your name please?

I: My name is Robert Triggs.

C: How many years have you been in Saudi Arabia?

I: I have been in Saudi Arabia for just over one year.

C: Where have you been before you came here?

I: Before that I was teaching in London for six years and I have taught in France and Spain.

C: Is there any difference between the life in France and Britain and the life in Saudi Arabia?

I: A lot of difference. The food here is different. Driving of course is different. In Saudi Arabia you drive on the right side of the road, we drive on the left. The language of course is different, even the dress is different.

C: Yes.

I: O.K. Well, that's fine. Thank you very much for coming.

Phonemic transcription of part of an interview

denoting level two

Interviewer: O.K. then Khalid. Well thank you for coming here today and first of all can you tell me a little bit about your home. What part of Saudi Arabia do you come from?

Cadet: I come from Riyadh.

I: You come from Riyadh, have your family always lived in Riyadh?

C: Yes all of them.

I: All of them, good. What about your education. Tell me your education, your primary school, your secondary school, your high school.

C: In Riyadh, all of them in Riyadh.

I: How old were you when you went to school first of all?

C: In primary. In my foot, in my feet. I go to school in the morning.

I: Yah, but how old were you?

C: In six years.

I: And how long were you at primary school?

C: Six years.

I: I was.

C: I was at primary school six year then secondary school three years then high school three years.

I: In Riyadh.

C: Yes in Riyadh.

I: What subjects did you study at high school?

C: High school, many subjects culture, scientific culture.

I: For example. Did you study physics, mathematics?

C: Yes, physics, mathematics, algebra, chemistry, biology.

I: When did you start learning English?

C: From primary school.

I: So you learnt English for ten years in school?

C: Yah.

I: Were the teachers from England or from Egypt or from Saudi Arabia.

C: No no. They were from Egypt and from Jordan.

I: And how many lessons of English did you have every week? Four hours a week?

C: Every week, no, seven hours a week.

I: Were the teachers good in English?

C: Yes most of them and other of them not good.

I: What kind of English did you do in school? Did you talk much English or was it mainly reading books and writing exercises.

C: Writing and reading books. Not talking.

I: And then when you left school. How old were you when you left school?

C: Twenty, twenty years.

I: Did you do any jobs, any temporary jobs during the holidays, during Ramadan or after you left school? What kind of jobs did you do?

C: I do in the many establish and company.

I: You were working for a short time?

C: Yes, short time four month.

I: What kind of work was it?

C: Like writer or go to another company or make decision.

I: What kind of company was it, a building company or a post office?

C: No no supply, supplyment.

I: Good. What kind of job does your father do?

C: He is an officer.

I: In the Air Force?

C: Yes in the Air Force.

I: In this Academy?

C: No no in the base.

I: In the base. I see. What is his rank?

C: His rank major.

I: Has he always been in the Air Force?

C: No no he was before in the Army.

I: Really and what rank was he in the Army?

C: Captain.

I: Is it easy to change from the Army to the Air Force?

C: Yah, it is not easy, but he sent an application to Prince Sultan the Minister of Aviation Defense and he approved the application.

I: And why did he change from being in the Army to being in the Air Force?

C: The Air Force in Riyadh, he was in Tabuk but the base in Riyadh and my family in Riyadh.

I: Tell us about your family. How big is your family?

C: It's a big family, five boys and three girls.

I: Are you the oldest?

C: No no I am the four.

I: The fourth.
Have you been outside Saudi Arabia, have you been abroad?

C: Yes.

I: Which countries have you visited?

C: I visited Jordan.

A full item analysis of discrete items tests

Listening and grammatical structure tests included in T.E.A.C.

I Listening Comprehension Test (lecturettes)

Item No.	Item Facility	Item Discrimination
1	+.36	+.40
2	+.50	+.52
3	*.60	+.08
4	+.60	+.16
5	+.54	+.48
6	+.52	+.40
7	+.74	+.12
8	+.70	+.28
9	+.50	+.44
10	+.64	+.24
11	+.64	+.32
12	+.56	+.16
13	+.56	+.32

II Grammatical Structure Test

1	+.62	+.35
2	+.58	+.26
3	+.64	+.25
4	+.87	+.30
5	+.58	+.26
6	+.48	+.27
7	+.19	+.13
8	+.56	+.30
9	+.72	+.56
10	+.76	+.36
11	+.59	+.36
12	+.30	+.30
13	+.74	+.36
14	+.32	+.38
15	+.58	+.51
16	+.20	+0.03
17	+.46	+.48
18	+.46	+.45
19	+.83	+.53
20	+.56	+.50
21	+.72	+.37
22	+.74	+.35
23	+.59	+.26
24	+.40	+.36
25	+.51	+.46
26	+.24	+.30
27	+.74	+.26

Item No.	Item Facility	Item Discrimination
28	+.61	+.46
29	+.82	+.33
30	+.75	+.26
31	+.37	+.25
32	+0.08	+0.03
33	+.14	+0.00
34	+.48	+.27
35	+.48	+.26
36	+0.08	+.03
37	+.19	+.20
38	+.59	+.30
39	+.50	+.31
40	+.59	+.46
41	+.24	+.10
42	+.64	+.40
43	+.22	+.16
44	+.82	+.50
45	+.74	+.50
46	+.58	+.33
47	+.16	+.13
48	+.38	+.26
49	+.58	+.36
50	+.69	+.30
51	+.40	+.27
52	+.37	+.33
53	+.27	+.30
54	+.82	+.36
55	+.35	+.32
56	+.64	+.33
57	+.69	+.58
58	+.70	+.25
59	+.72	+.41
60	+.91	+.13
61	+.25	+.26
62	+.32	+.40
63	+.75	+.37
64	+.53	+.36
65	+.62	+.35
66	+.24	+.30
67	+.38	+.22
68	+.40	+.25
69	+.87	+.20
70	+.64	+.51
71	+.46	+.41
72	+.25	+.13
73	+.58	+.38

T.E.A.C.

Scale Standard Scores
Mean = 5 SD = 2 N = 92 level

Service No.	Gram-mar	List-ening	Dict-ation 1	Dict-ation 2	Cloze 1	Cloze 2	Cloze 3	Cloze 4	Cloze total	Oral inter-view	T.E.A.C total
1000	27	3	29.5	31.5	3	1.5	3	1.5	16.5	5	120
1639	23	3	28	29	2	0.5	3.5	3	16.5	5	112
1711	22.5	3	26.5	28.5	2.5	2	4	3.5	19.5	5	112.5
1708	22	2.5	24	24	2.5	2	2.5	2	16.5	4	101
1741	22	2.5	24.5	28	0.5	2	3	3	16	5	105.5
1549	22	2.5	27	29	1.5	1.5	2.5	2	15	5	107.5
1571	21.5	2.5	26.5	27	0	1.5	2	3.5	14.5	4	99
1631	21.5	2.5	24	28	2	1.5	3	2.5	16.5	5	105
1700	21	2.5	23.5	29.5	2.5	0.5	2.5	2.5	15.5	5	104.5
1663	21	2.5	26	29	1.5	1.5	2	2.5	15	4	105
1590	21	2.5	26.5	28.5	1	1.5	2	1	13	4	103.5
1344	21	2.5	25	28	1	1.5	2	1.5	13.5	5	102.5
1467	21	2.5	24.5	29.5	1.5	2.5	3	1.5	16	5	105.5
1577	21	2	27	29.5	1	1	2	1	12.5	4	101.5
1512	20.5	2	24	28.5	-1	1	0	1.5	8.5	4	95.5
1521	20.5	2	23.5	28	-0.5	1.5	1	-1	8.5	4	94.5
1558	20.5	2	24.5	29	2.5	2.5	2.5	2.5	17.5	5	106
1618	20.5	2	26	31.5	1.5	0	2.5	1	12.5	4	104.5
1569	20.5	2	23.5	30	1	0.5	2	1.5	12.5	5	101
1673	20.5	2	30	31	0.5	1	2	2.5	13.5	5	109
1583	20	2	26.5	29.5	1.5	1	2	0	12	4	104
1372	20	2	28	31	0	1.5	1.5	2	12.5	4	106
1702	20	2	29	30.5	2	2.5	3	3	18	5	112.5
1691	19.5	2	29.5	31.5	1.5	2	2	2	15	5	110
1559	19.5	2	24.5	25.5	0.5	2	1.5	1.5	13	4	96.5
1599	19	2	26.5	30	2	2.5	2.5	2.5	17	5	107
1528	19	2	26	28	-1	0.5	2.5	2	11.5	4	98.5
1531	19	2	24.5	26	0.5	0.5	2	2	12.5	4	96
1592	19	2	28.5	29.5	-0.5	0.5	1	1.5	10	4	101
1555	18.5	2	26.5	27	1.5	2	3	2	16	4	102
1617	18.5	1.5	29	29.5	-1	2	1.5	2	12	4	102.5
1540	18.5	1.5	28.5	31.5	-0.5	-0.5	0.5	2	9	4	101
1565	18.5	1.5	27	29.5	0.5	-0.5	2	1.5	10	4	98.5
1542	18.5	1.5	25.5	29.5	-0.5	1	1	1	10	4	97
1543	18.5	1.5	24.5	26	-1.5	1.5	1.5	0.5	9.5	4	92
1609	18	1.5	24.5	25.5	0	0.5	1.5	2	11.5	4	93
1581	18	1.5	27	29.5	2.5	0.5	2	1	13.5	4	101.5
1478	18	1.5	24	27	1.5	1.5	1	0.5	12	4	94.5
1493	17.5	1.5	25.5	28.5	1	1.5	1.5	2	13.5	4	98.5
1603	17.5	1.5	23	26.5	-0.5	0	0	0	7	4	87.5
1611	17.5	1	27.5	30	0.5	0.5	1.5	2	12	3	99.5
1548	17.5	1	24	27	-1	-0.5	2	1.5	9.5	3	90.5

T.E.A.C.

Scale Standard Scores
Mean = 5 SD = 2 N = 92 level

Service No.	Gram-mar	List-ening	Dict-ation 1	Dict-ation 2	Cloze 1	Cloze 2	Cloze 3	Cloze 4	Cloze total	Oral inter-	T.E.A.C total
1466	17.5	1	23.5	27	2.5	2	2.5	1	15.5	4	96.5
1551	17.5	1	24.5	26.5	2.5	0	2.5	-0.5	13	4	94.5
1537	17.5	1	23	27	1.5	-1	0.5	1	9.5	3	89.5
1529	17	1	23.5	26.5	-0.5	0.5	3	1	11.5	3	91
1645	17	1	21.5	22.5	-1	-0.5	0.5	1.5	8	3	81.5
1560	17	1	21	23.5	1.5	2	4	2	17	4	91.5
1534	17	1	19.5	23.5	0.5	1.5	2	2	13.5	3	86
1498	16.5	1	21.5	24	-1	0	1	0.5	8	3	82.5
1624	16.5	1	21	23.5	-1.5	2	1	-1	8	3	81.5
1519	16	0.5	19.5	22.5	-1.5	0	1.5	1	8.5	3	78.5
1347	16	0.5	17.5	21.5	-1	-0.5	0	1.5	7.5	3	74
1481	16	0.5	18	21	0	1	0	0	8.5	3	75.5
1527	16	0.5	19	22	-2	1	2	-0.5	8	3	77
1576	16	0.5	19	19.5	-2	-0.5	1.5	-0.5	6	3	72.5
1656	15.5	0.5	17.5	21.5	-0.5	0.5	0	-0.5	7	3	73.5
1550	15.5	0.5	18.5	19.5	0.5	-1	0	1	8	3	73.5
1518	15.5	0.5	18	21.5	-0.5	0.5	-0.5	0	7	3	74
1586	15.5	0.5	23.5	26	-1.5	-0.5	1	-0.5	6	2	82.5
1579	15.5	0.5	23.5	27	-2	1	-0.5	0.5	6.5	2	84
1600	15	0.5	20.5	23.5	-0.5	-1.5	-1	1	5.5	2	76
1561	15	0	20.5	24	0	0	-1	0.5	2	2	72.5
1587	15	0.5	15.5	19	-1	0	0	0	1.5	2	62.5
1616	14	0.5	14	19.5	-1	-0.5	-2	0	1.5	2	61.5
1608	14.5	0	14	18	-1	0	-0.5	0	6	2	63.5
1710	14.5	0.5	15.5	20	0	-0.5	1.5	0.5	10	2	71.5
1546	14	0.5	15	21.5	-2	-1	-1	0.5	4	2	66
1552	13.5	0.5	16	22	-1	-1.5	-0.5	0	4.5	2	67.5
1337	13.5	0.5	15	18	-2	-1	-1	-0.5	3	2	61
1532	13.5	0	19	18	-1	-0.5	0	0	6	2	67.5
1669	13.5	0	21.5	21	-1	0	0.5	-0.5	6.5	2	73.5
1516	14	0	21	20	0	-1	-1	-2	3.5	2	69.5
1567	14	0.5	24	25	-1	0	1	1.5	9	2	84.5
1500	13	0	15.5	17	1.5	0.5	2	1.5	13	2	74.5
1596	13.5	0	14	16	-1.5	-0.5	0	-1	4.5	2	59
1545	12.5	0	13.5	17	1.5	0.5	2	1.5	13	3	67.5
1525	13	0.5	15.5	17	0	-0.5	-1	-1	2.5	2	59.5
1484	13	0.5	16	17.5	0.5	0.5	3	0	11.5	2	69.5
1629	13	0.5	19	19.5	-1	0	-0.5	-1	5	2	68
1446	12.5	0	25	25	-0.5	0.5	0.5	-1.5	6.5	2	80
1570	12	0	23.5	24	-2	-2.5	-0.5	-1	1.5	2	72

T.E.A.C.

Scale Standard Scores
Mean = 5 SD = 2 N = 92 level

Service No.	Gram-mar	List-ening	Dict-ation 1	Dict-ation 2	Cloze 1	Cloze 2	Cloze 3	Cloze 4	Cloze total	Oral inter-	T.E.A.C total
1491	12	0	14	16	2	0	1.5	0	11	2	64
1564	12	0	13	16.5	-1.5	-0.5	-1	-2.5	2	2	54.5
1447	11.5	0	14.5	17	0.5	1	1	-2.5	7.5	3	62
1597	11	0	13.5	13	-0.5	-1.5	0.5	-2.5	3.5	3	52
1706	10.5	0	15.5	16	-2.5	-1.5	-2	-2.5	3.5	2	52
1568	10	0	15.5	16	0.5	1.5	1.5	0.5	11.5	2	64
1524	10	-0.5	15	17	-1	0.5	0.5	-0.5	7	2	59.5
1386	10	-0.5	14.5	15.5	0	1	2	1.5	12	2	62.5
1585	9	-0.5	16	14.5	-1	-0.5	-1.5	-1.5	3	2	51.5
1684	6.5	-1	9.5	13.5	-1	-1.5	-1	-2.5	1.5	1	40

TEST OF SPEAKING

Additional Comments by interviewers On Cadets' Performances on the Test

Cadet — Khalid is a confident, effective communicator. With practice he will easily reach level 5 and above. (N.B.) the cadet was given 4 points on the scale)

Cadet — Although quietly spoken, Hamid communicates effectively, in short, carefully constructed sentences. With more confidence and practice, he will reach level 5. (N.B. the cadet was given level 4)

Cadet — Sultan had difficulty in forming grammatical utterances and hesitated frequently during his responses. He would usually "dry up" after 2 or 3 sentences. Main problem seems to be an inadequate grasp of grammatical structure. (N.B. the cadet was given level 2)

Cadet — Malki is able to construct simple sentences effectively when he gives his speech full attention. Many of his grammatical and lexical errors are caused by carelessness rather than ignorance. (N.B. the cadet was given level 3)

- - - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - -

	DYN	AIE	AVN	AIR	NAV	MET	AV
ALCPT	-0.1399	-0.0682	0.0106	0.0926	0.1130	0.1122	0.0242
	(63)	(63)	(63)	(63)	(63)	(63)	(63)
	P=0.137	P=0.298	P=0.467	P=0.235	P=0.189	P=0.191	P=0.425

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 2

24 JAN 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

VARIABLE	CASES	MEAN	STD DEV
GR	92	38.4891	7.4620
LIS	92	7.2174	1.9320
DIC1	92	43.5217	9.9388
DIC2	92	53.4565	10.2661
CL1	92	5.1522	2.8437
CL2	92	5.9891	2.2943
CL3	92	7.2826	2.8721
CL4	92	6.4239	2.9990
CLT	92	24.8261	9.3991
SP	92	3.2826	1.1225
TEACT	92	175.7283	36.4725

24 JAN 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

TABLE 3

03 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

VARIABLE	CASES	MEAN	STD DEV
GR	47	44.4468	3.9166
LIS	47	8.8298	1.1290
DIC1	47	56.4681	4.0906
DIC2	47	61.8085	3.9817
CL1	47	6.7234	2.4557
CL2	47	7.1489	1.8295
CL3	47	8.9362	1.7986
CL4	47	8.3617	1.7622
CLT	47	31.1277	5.9000
SP	47	4.2128	.6233
TEACT	47	206.6170	14.5953

TABLE 4

03 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

VARIABLE	CASES	MEAN	STD DEV
GR	45	32.2667	4.6486
LIS	45	5.5333	.8421
DIC1	45	40.2222	6.9931
DIC2	45	44.7333	7.0045
CL1	45	3.5111	2.2424
CL2	45	4.7778	2.1094
CL3	45	5.5556	2.7762
CL4	45	4.4000	2.6748
CLT	45	18.2444	7.7200
SP	45	2.3111	.5569
TEACT	45	143.4667	21.0914

TABLE 5

15 JAN 86 COPMUTING DATA
University of Edinburgh - ERCC ICL 2976 EMAS

VARIABLE	CASES	MEAN	STD DEV
GR	8	50.2250	27.6203
LIS	8	74.0125	10.0290
DIC1	8	41.2125	26.2667
DIC2	8	17.0750	22.3777
CL1	8	47.4625	35.5078
CL2	8	14.2375	21.1768
CL3	8	71.4750	26.0727
CL4	8	63.1500	34.6223
SP	8	4.2500	.7071

TABLE 6

- - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - - 86.

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
	1.0000	0.8686	0.5518	0.7247	-0.0548	0.2542	0.2031	0.4390	0.2787	0.6520
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=*****	P=0.000	P=0.000	P=0.000	P=0.360	P=0.046	P=0.090	P=0.001	P=0.032	P=0.000
	0.8686	1.0000	0.4618	0.6373	-0.0273	0.3113	0.2204	0.3572	0.2802	0.6075
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=*****	P=0.001	P=0.000	P=0.429	P=0.019	P=0.073	P=0.008	P=0.031	P=0.000
1	0.5518	0.4618	1.0000	0.8633	-0.1799	0.0851	0.1515	0.2795	0.1223	0.2678
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=0.001	P=*****	P=0.000	P=0.118	P=0.289	P=0.160	P=0.031	P=0.212	P=0.038
2	0.7247	0.6373	0.8633	1.0000	-0.2169	0.1113	0.1048	0.3879	0.1395	0.3422
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=0.000	P=0.000	P=*****	P=0.076	P=0.233	P=0.247	P=0.004	P=0.180	P=0.011
	-0.0548	-0.0273	-0.1799	-0.2169	1.0000	0.4666	0.5703	0.3592	0.7475	0.2702
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.360	P=0.429	P=0.118	P=0.076	P=*****	P=0.001	P=0.000	P=0.008	P=0.000	P=0.036
	0.2542	0.3113	0.0851	0.1113	0.4666	1.0000	0.6192	0.3182	0.7417	0.5052
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.046	P=0.019	P=0.289	P=0.233	P=0.001	P=*****	P=0.000	P=0.017	P=0.000	P=0.000
	0.2031	0.2204	0.1515	0.1048	0.5703	0.6192	1.0000	0.5785	0.8949	0.4884
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.090	P=0.073	P=0.168	P=0.247	P=0.000	P=0.000	P=*****	P=0.000	P=0.000	P=0.000
	0.4390	0.3572	0.2795	0.3879	0.3592	0.3182	0.5785	1.0000	0.7458	0.4028
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.001	P=0.008	P=0.031	P=0.004	P=0.008	P=0.017	P=0.000	P=*****	P=0.000	P=0.003
	0.2787	0.2802	0.1223	0.1395	0.7475	0.7417	0.8949	0.7458	1.0000	0.5317
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.032	P=0.031	P=0.212	P=0.180	P=0.000	P=0.000	P=0.000	P=0.000	P=*****	P=0.000
	0.6520	0.6075	0.2678	0.3422	0.2702	0.5052	0.4884	0.4028	0.5317	1.0000
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=0.000	P=0.038	P=0.011	P=0.036	P=0.000	P=0.000	P=0.003	P=0.000	P=*****
ACT	0.8040	0.7176	0.8116	0.8644	0.1582	0.4279	0.4989	0.6368	0.5629	0.5872
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=0.000	P=0.000	P=0.000	P=0.150	P=0.002	P=0.000	P=0.000	P=0.000	P=0.000

TABLE 7

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
GR	1.0000 (47) P=*****	0.9074 (47) P=0.000	0.3191 (47) P=0.014	0.3444 (47) P=0.009	0.4290 (47) P=0.001	0.3940 (47) P=0.003	0.4825 (47) P=0.000	0.3729 (47) P=0.005	0.5638 (47) P=0.000	0.6637 (47) P=0.000
LIS	0.9074 (47) P=0.000	1.0000 (47) P=*****	0.3000 (47) P=0.020	0.2876 (47) P=0.025	0.3982 (47) P=0.003	0.4862 (47) P=0.000	0.5084 (47) P=0.000	0.4796 (47) P=0.000	0.6201 (47) P=0.000	0.7322 (47) P=0.000
DIC1	0.3191 (47) P=0.014	0.3000 (47) P=0.020	1.0000 (47) P=*****	0.7237 (47) P=0.000	0.1170 (47) P=0.217	0.2142 (47) P=0.074	0.2287 (47) P=0.061	0.3500 (47) P=0.008	0.2848 (47) P=0.026	0.3182 (47) P=0.015
DIC2	0.3444 (47) P=0.009	0.2876 (47) P=0.025	0.7237 (47) P=0.000	1.0000 (47) P=*****	0.2212 (47) P=0.067	0.1264 (47) P=0.199	0.1682 (47) P=0.129	0.1743 (47) P=0.121	0.2306 (47) P=0.059	0.4547 (47) P=0.001
CL1	0.4290 (47) P=0.001	0.3982 (47) P=0.003	0.1170 (47) P=0.217	0.2212 (47) P=0.067	1.0000 (47) P=*****	0.3529 (47) P=0.007	0.6111 (47) P=0.000	0.2246 (47) P=0.065	0.7812 (47) P=0.000	0.5080 (47) P=0.000
CL2	0.3940 (47) P=0.003	0.4862 (47) P=0.000	0.2142 (47) P=0.074	0.1264 (47) P=0.199	0.3529 (47) P=0.007	1.0000 (47) P=*****	0.4390 (47) P=0.001	0.3133 (47) P=0.016	0.6971 (47) P=0.000	0.5626 (47) P=0.000
CL3	0.4825 (47) P=0.000	0.5084 (47) P=0.000	0.2287 (47) P=0.061	0.1682 (47) P=0.129	0.6111 (47) P=0.000	0.4390 (47) P=0.001	1.0000 (47) P=*****	0.5081 (47) P=0.000	0.8469 (47) P=0.000	0.5165 (47) P=0.000
CL4	0.3729 (47) P=0.005	0.4796 (47) P=0.000	0.3500 (47) P=0.008	0.1743 (47) P=0.121	0.2246 (47) P=0.065	0.3133 (47) P=0.016	0.5081 (47) P=0.000	1.0000 (47) P=*****	0.6457 (47) P=0.000	0.4430 (47) P=0.001
CLT	0.5638 (47) P=0.000	0.6201 (47) P=0.000	0.2848 (47) P=0.026	0.2306 (47) P=0.059	0.7812 (47) P=0.000	0.6971 (47) P=0.000	0.8469 (47) P=0.000	0.6457 (47) P=0.000	1.0000 (47) P=*****	0.6782 (47) P=0.000
SP	0.6637 (47) P=0.000	0.7322 (47) P=0.000	0.3182 (47) P=0.015	0.4547 (47) P=0.001	0.5080 (47) P=0.000	0.5626 (47) P=0.000	0.5165 (47) P=0.000	0.4430 (47) P=0.001	0.6782 (47) P=0.000	1.0000 (47) P=*****
TEACT	0.7568 (47) P=0.000	0.7453 (47) P=0.000	0.7080 (47) P=0.000	0.7124 (47) P=0.000	0.5877 (47) P=0.000	0.5363 (47) P=0.000	0.6408 (47) P=0.000	0.5262 (47) P=0.000	0.7658 (47) P=0.000	0.7666 (47) P=0.000

TABLE 8

- - - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - -

	TEACT
GR	0.7568 (47) P=0.000
LIS	0.7453 (47) P=0.000
DIC1	0.7080 (47) P=0.000
DIC2	0.7124 (47) P=0.000
CL1	0.5877 (47) P=0.000
CL2	0.5363 (47) P=0.000
CL3	0.6408 (47) P=0.000
CL4	0.5262 (47) P=0.000
CLT	0.7658 (47) P=0.000
SP	0.7666 (47) P=0.000
TEACT	1.0000 (47) P=*****

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A [illegible] CANNOT BE COMPUTED)

- - - - - - - - - - - - PEARSON CORRELATION COEFFICIENTS - - - - - - - - - - - -

TABLE 9

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	SP
GR	1.0000 (8) P=*****	0.6067 (8) P=0.061	-0.0105 (8) P=0.491	-0.3732 (8) P=0.190	0.0420 (8) P=0.462	0.2021 (8) P=0.322	-0.2013 (8) P=0.323	-0.2179 (8) P=0.309	0.7084 (8) P=0.028
LIS	0.6067 (8) P=0.061	1.0000 (8) P=*****	-0.0040 (8) P=0.496	-0.0690 (8) P=0.438	-0.2536 (8) P=0.280	0.1451 (8) P=0.371	-0.2855 (8) P=0.255	-0.0955 (8) P=0.414	0.8919 (8) P=0.002
DIC1	-0.0105 (8) P=0.491	-0.0040 (8) P=0.496	1.0000 (8) P=*****	-0.1591 (8) P=0.359	0.8658 (8) P=0.003	0.3923 (8) P=0.177	0.4722 (8) P=0.127	0.8051 (8) P=0.009	0.0267 (8) P=0.476
DIC2	-0.3732 (8) P=0.190	-0.0690 (8) P=0.438	-0.1591 (8) P=0.359	1.0000 (8) P=*****	-0.2807 (8) P=0.258	-0.2288 (8) P=0.300	0.3944 (8) P=0.176	0.4344 (8) P=0.150	-0.2361 (8) P=0.294
CL1	0.0420 (8) P=0.462	-0.2536 (8) P=0.280	0.8658 (8) P=0.003	-0.2807 (8) P=0.258	1.0000 (8) P=*****	0.2097 (8) P=0.316	0.3610 (8) P=0.199	0.6621 (8) P=0.041	-0.1156 (8) P=0.397
CL2	0.2021 (8) P=0.322	0.1451 (8) P=0.371	0.3923 (8) P=0.177	-0.2288 (8) P=0.300	0.2097 (8) P=0.316	1.0000 (8) P=*****	0.0704 (8) P=0.437	0.2902 (8) P=0.251	0.4524 (8) P=0.139
CL3	-0.2013 (8) P=0.323	-0.2855 (8) P=0.255	0.4722 (8) P=0.127	0.3944 (8) P=0.176	0.3610 (8) P=0.199	0.0704 (8) P=0.437	1.0000 (8) P=*****	0.6643 (8) P=0.041	-0.3220 (8) P=0.227
CL4	-0.2179 (8) P=0.309	-0.0955 (8) P=0.414	0.8051 (8) P=0.009	0.4344 (8) P=0.150	0.6621 (8) P=0.041	0.2902 (8) P=0.251	0.6643 (8) P=0.041	1.0000 (8) P=*****	-0.0957 (8) P=0.414
SP	0.7084 (8) P=0.028	0.8919 (8) P=0.002	0.0267 (8) P=0.476	-0.2361 (8) P=0.294	-0.1156 (8) P=0.397	0.4524 (8) P=0.139	-0.3220 (8) P=0.227	-0.0957 (8) P=0.414	1.0000 (8) P=*****

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 10

03 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - -

	TEACT
GR	0.8040 (45) P=0.000
LIS	0.7176 (45) P=0.000
DIC1	0.8116 (45) P=0.000
DIC2	0.8644 (45) P=0.000
CL1	0.1582 (45) P=0.150
CL2	0.4279 (45) P=0.002
CL3	0.4989 (45) P=0.000
CL4	0.6368 (45) P=0.000
CLT	0.5629 (45) P=0.000
SP	0.5872 (45) P=0.000
TEACT	1.0000 (45) P=*****

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

03 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - - - S P E A R M A N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - -

TABLE 11

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
LIS	.9565 N(47) SIG .000									
DIC1	.2647 N(47) SIG .036	.2613 N(47) SIG .038								
DIC2	.2817 N(47) SIG .028	.2203 N(47) SIG .068	.6838 N(47) SIG .000							
CL1	.3735 N(47) SIG .005	.3913 N(47) SIG .003	.0967 N(47) SIG .259	.2034 N(47) SIG .085						
CL2	.4191 N(47) SIG .002	.4610 N(47) SIG .001	.1972 N(47) SIG .092	.0737 N(47) SIG .311	.3624 N(47) SIG .006					
CL3	.4838 N(47) SIG .000	.5183 N(47) SIG .000	.1613 N(47) SIG .139	.1106 N(47) SIG .238	.6317 N(47) SIG .000	.4005 N(47) SIG .003				
CL4	.4476 N(47) SIG .001	.4969 N(47) SIG .000	.3582 N(47) SIG .007	.2039 N(47) SIG .085	.2070 N(47) SIG .081	.3557 N(47) SIG .007	.4350 N(47) SIG .001			
CLT	.5921 N(47) SIG .000	.6423 N(47) SIG .000	.2785 N(47) SIG .029	.2100 N(47) SIG .078	.7964 N(47) SIG .000	.6931 N(47) SIG .000	.8028 N(47) SIG .000	.6154 N(47) SIG .000		
SP	.6950 N(47) SIG .000	.7060 N(47) SIG .000	.2821 N(47) SIG .027	.4167 N(47) SIG .002	.5124 N(47) SIG .000	.5219 N(47) SIG .000	.5501 N(47) SIG .000	.4973 N(47) SIG .000	.7144 N(47) SIG .000	
TEACT	.7132 N(47) SIG .000	.7206 N(47) SIG .000	.6500 N(47) SIG .000	.6772 N(47) SIG .000	.5557 N(47) SIG .000	.5361 N(47) SIG .000	.6076 N(47) SIG .000	.5913 N(47) SIG .000	.7736 N(47) SIG .000	.7823 N(47) SIG .000

- - - - - - - - - - - - - - - S P E A R M A N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - - -

TABLE 12

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
LIS	.8231 N(45) SIG .000									
DIC1	.5461 N(45) SIG .000	.4508 N(45) SIG .001								
DIC2	.7121 N(45) SIG .000	.6272 N(45) SIG .000	.8448 N(45) SIG .000							
CL1	-.0926 N(45) SIG .273	-.0686 N(45) SIG .327	-.1569 N(45) SIG .152	-.2124 N(45) SIG .081						
CL2	.2353 N(45) SIG .060	.2387 N(45) SIG .057	.1032 N(45) SIG .250	.1056 N(45) SIG .245	.4552 N(45) SIG .001					
CL3	.1847 N(45) SIG .112	.1571 N(45) SIG .151	.1300 N(45) SIG .197	.0867 N(45) SIG .286	.5008 N(45) SIG .000	.6546 N(45) SIG .000				
CL4	.4151 N(45) SIG .002	.3281 N(45) SIG .014	.2448 N(45) SIG .053	.3494 N(45) SIG .009	.3466 N(45) SIG .010	.3168 N(45) SIG .017	.5320 N(45) SIG .000			
CLT	.2866 N(45) SIG .028	.2585 N(45) SIG .043	.1418 N(45) SIG .176	.1396 N(45) SIG .188	.6568 N(45) SIG .000	.7672 N(45) SIG .000	.8854 N(45) SIG .000	.7238 N(45) SIG .000		
SP	.6629 N(45) SIG .000	.5488 N(45) SIG .000	.2511 N(45) SIG .048	.3298 N(45) SIG .014	.2180 N(45) SIG .075	.4673 N(45) SIG .001	.4503 N(45) SIG .001	.3548 N(45) SIG .009	.5306 N(45) SIG .000	
TEACT	.7589 N(45) SIG .000	.6351 N(45) SIG .000	.8129 N(45) SIG .000	.8540 N(45) SIG .000	.1347 N(45) SIG .189	.4120 N(45) SIG .002	.4678 N(45) SIG .001	.5865 N(45) SIG .000	.5542 N(45) SIG .000	.5294 N(45) SIG .000

TABLE 13

ANALYSIS NUMBER 1 LISTWISE DELETION OF CASES WITH MISSING VALUES

EXTRACTION 1 FOR ANALYSIS 1, PRINCIPAL-COMPONENTS ANALYSIS (PC)

INITIAL STATISTICS:

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	1.00000	1	4.07161	45.2	45.2
LIS	1.00000	2	2.26210	25.1	70.4
DIC1	1.00000	3	.86295	9.6	80.0
DIC2	1.00000	4	.61148	6.8	86.8
CL1	1.00000	5	.39895	4.4	91.2
CL2	1.00000	6	.35439	3.9	95.1
CL3	1.00000	7	.24868	2.8	97.9
CL4	1.00000	8	.10419	1.2	99.0
SP	1.00000	9	.08565	1.0	100.0

PC EXTRACTED 2 FACTORS.

FACTOR MATRIX:

	FACTOR 1	FACTOR 2
GR	.85909	-.31704
LIS	.81730	
SP	.76895	
DIC2	.74596	-.54501
CL4	.66062	
DIC1	.64426	-.49801
CL1		.81736
CL3	.56946	.65828
CL2	.54194	.57052

FINAL STATISTICS:

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	.83854	1	4.07161	45.2	45.2
LIS	.73439	2	2.26210	25.1	70.4
DIC1	.66308				
DIC2	.85349				
CL1	.71647				
CL2	.61919				
CL3	.75762				
CL4	.51670				
SP	.63423				

ANALYSIS NUMBER 1 LISTWISE DELETION OF CASES WITH MISSING VALUES

EXTRACTION 1 FOR ANALYSIS 1. PRINCIPAL-COMPONENTS ANALYSIS (PC)

INITIAL STATISTICS:

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	1.00000	1	4.37911	48.7	48.7
LIS	1.00000	2	1.41138	15.7	64.3
DIC1	1.00000	3	.82726	9.2	73.5
DIC2	1.00000	4	.81554	9.1	82.6
CL1	1.00000	5	.67209	7.5	90.1
CL2	1.00000	6	.38214	4.2	94.3
CL3	1.00000	7	.27839	3.1	97.4
CL4	1.00000	8	.16277	1.8	99.2
SP	1.00000	9	.07131	.8	100.0

PC EXTRACTED 2 FACTORS.

FACTOR MATRIX:

	FACTOR 1	FACTOR 2
LIS	.85793	
SP	.85669	
GR	.82464	
CL3	.72597	-.30471
CL2	.63097	
CL1	.62441	-.30198
CL4	.61468	
DIC2	.51595	.75888
DIC1	.51964	.75215

FINAL STATISTICS:

TABLE 15

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	.62888	1	5.17463	51.7	51.7
LIS	.68654	2	1.50454	15.0	66.8
DIC1	.79480				
DIC2	.82372				
CL1	.54567				
CL2	.48251				
CL3	.68210				
CL4	.40324				
CLT	.92328				
SP	.70851				

FACTOR MATRIX: TABLE 16

	FACTOR 1	FACTOR 2
CLT	.91480	
SP	.83935	
LIS	.82838	
GR	.78870	
CL3	.77139	
CL1	.67398	-.30236
CL2	.66076	
CL4	.63487	
DIC2	.45959	.78263
DIC1	.47416	.75497

INITIAL STATISTICS:

TABLE 17

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	1.00000	1	4.61927	46.2	46.2
LIS	1.00000	2	2.67929	26.8	73.0
DIC1	1.00000	3	.89379	8.9	81.9
DIC2	1.00000	4	.61163	6.1	88.0
CL1	1.00000	5	.40013	4.0	92.0
CL2	1.00000	6	.35724	3.6	95.6
CL3	1.00000	7	.24879	2.5	98.1
CL4	1.00000	8	.10420	1.0	99.1
CLT	1.00000	9	.08565	.9	100.0
SP	1.00000	10	-.00000	-.0	100.0

PC EXTRACTED 2 FACTORS.

FACTOR MATRIX:

TABLE 18

	FACTOR 1	FACTOR 2
CLT	.81953	-.55241
SP	.77093	
GR	.73974	.53988
CL4	.73422	
CL3	.72731	-.50349
LIS	.71094	.47705
CL2	.66240	-.40287
CL1	.42207	-.73581
DIC2	.59471	.69986
DIC1	.51185	.62314

TABLE 19

ANALYSIS NUMBER 1 LISTWISE DELETION OF CASES WITH MISSING VALUES

EXTRACTION 1 FOR ANALYSIS 1, PRINCIPAL-COMPONENTS ANALYSIS (PC)

INITIAL STATISTICS:

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	1.00000	1	5.17463	51.7	51.7
LIS	1.00000	2	1.50454	15.0	66.8
DIC1	1.00000	3	.92585	9.3	76.1
DIC2	1.00000	4	.81647	8.2	84.2
CL1	1.00000	5	.67528	6.8	91.0
CL2	1.00000	6	.38224	3.8	94.8
CL3	1.00000	7	.28311	2.8	97.6
CL4	1.00000	8	.16509	1.7	99.3
CLT	1.00000	9	.07132	.7	100.0
SP	1.00000	10	.00147	.0	100.0

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER NORMALIZATION.

VARIMAX CONVERGED IN 3 ITERATIONS. TABLE 20

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2
DIC2	.92336	
QR	.88836	
LIS	.82040	
DIC1	.81286	
CL3		.86471
CL1		.79920
CL2		.77670
CL4	.38610	.60632
SP	.51846	.60450

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2
FACTOR 1	.82612	.56350
FACTOR 2	-.56350	.82612

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER NORMALIZATION.

TABLE 21

VARIMAX CONVERGED IN 3 ITERATIONS.

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2
CLT	.95596	
CL3	.82499	
LIS	.75183	.34827
SP	.74375	.39413
CL1	.73866	
CL2	.69126	
GR	.68978	.39115
CL4	.58717	
DIC2		.90118
DIC1		.88167

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2
FACTOR 1	.91630	.40049
FACTOR 2	-.40049	.91630

FINAL STATISTICS: TABLE 22

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	.68438	1	4.37911	48.7	48.7
LIS	.75293	2	1.41138	15.7	64.3
DIC1	.83575				
DIC2	.84210				
CL1	.48107				
CL2	.46043				
CL3	.61988				
CL4	.37843				
SP	.73552				

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER NORMALIZATION.

VARIMAX CONVERGED IN 3 ITERATIONS.

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2
LIS	.83089	
SP	.79127	.33077
CL3	.78652	
GR	.77341	
CL1	.69357	
CL2	.67707	
CL4	.56586	
DIC2		.90674
DIC1		.90224

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2
FACTOR 1	.90361	.42836
FACTOR 2	-.42836	.90361

FINAL STATISTICS:

TABLE 23

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
GR	.83869	1	4.61927	46.2	46.2
LIS	.73301	2	2.67929	26.8	73.0
DIC1	.65030				
DIC2	.84348				
CL1	.71955				
CL2	.60108				
CL3	.78248				
CL4	.55791				
CLT	.97678				
SP	.59526				

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER NORMALIZATION.

VARIMAX CONVERGED IN 3 ITERATIONS.

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2
CLT	.97946	
CL3	.87804	
CL1	.80444	
CL2	.76266	
CL4	.63973	.38555
SP	.55577	.53515
DIC2		.91818
GR		.89507
LIS		.82898
DIC1		.80578

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2
FACTOR 1	.74707	.66475
FACTOR 2	-.66475	.74707

TABLE 24 * * * A N A L Y S I S O F V A R I A N C E * * *

BY CL4 LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	360.816	1	360.816	70.957	0.000
LEVEL	360.816	1	360.816	70.957	0.000
EXPLAINED	360.816	1	360.816	70.957	0.000
RESIDUAL	457.651	90	5.085		
TOTAL	818.467	91	8.994		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 25 * * * A N A L Y S I S O F V A R I A N C E * * *

BY SP LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	83.135	1	83.135	237.403	0.000
LEVEL	83.135	1	83.135	237.403	0.000
EXPLAINED	83.135	1	83.135	237.403	0.000
RESIDUAL	31.517	90	0.350		
TOTAL	114.652	91	1.260		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 26 * * * A N A L Y S I S O F V A R I A N C E * * *

BY CLT LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	3815.672	1	3815.672	81.309	0.000
LEVEL	3815.672	1	3815.672	81.309	0.000
EXPLAINED	3815.672	1	3815.672	81.309	0.000
RESIDUAL	4223.545	90	46.928		
TOTAL	8039.217	91	88.343		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

27 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - Bush ICL 2988 EMAS

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

BY CLT LEVEL

TABLE 27 * * * A N A L Y S I S O F V A R I A N C E * * *

BY TEACT LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	91679.900	1	91679.900	280.917	0.000
LEVEL	91679.900	1	91679.900	280.917	0.000
EXPLAINED	91679.900	1	91679.900	280.917	0.000
RESIDUAL	29372.306	90	326.359		
TOTAL	121052.207	91	1330.244		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 28 * * * A N A L Y S I S O F V A R I A N C E * * *

BY DIC2 LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	6702.749	1	6702.749	208.875	0.000
LEVEL	6702.749	1	6702.749	208.875	0.000
EXPLAINED	6702.749	1	6702.749	208.875	0.000
RESIDUAL	2888.077	90	32.090		
TOTAL	9590.826	91	105.394		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 29 * * * A N A L Y S I S O F V A R I A N C E * * *

BY CL1 LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	237.221	1	237.221	42.815	0.000
LEVEL	237.221	1	237.221	42.815	0.000
EXPLAINED	237.221	1	237.221	42.815	0.000
RESIDUAL	498.649	90	5.541		
TOTAL	735.870	91	8.086		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 30 * * * A N A L Y S I S O F V A R I A N C E * * *

BY CL3 LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	262.733	1	262.733	48.463	0.000
LEVEL	262.733	1	262.733	48.463	0.000
EXPLAINED	262.733	1	262.733	48.463	0.000
RESIDUAL	487.920	90	5.421		
TOTAL	750.652	91	8.249		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 31 *** ANALYSIS OF VARIANCE ***

BY LIS
LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	249.814	1	249.814	250.264	0.000
LEVEL	249.814	1	249.814	250.264	0.000
EXPLAINED	249.814	1	249.814	250.264	0.000
RESIDUAL	89.838	90	0.998		
TOTAL	339.652	91	3.732		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

TABLE 32 * * * A N A L Y S I S O F V A R I A N C E * * *

DIC1
BY LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	6067.477	1	6067.477	186.917	0.000
LEVEL	6067.477	1	6067.477	186.917	0.000
EXPLAINED	6067.477	1	6067.477	186.917	0.000
RESIDUAL	2921.480	90	32.461		
TOTAL	8988.957	91	98.780		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

27 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - Bush ICL 2988 EMAS

TABLE 33 * * * A N A L Y S I S O F V A R I A N C E * * *

BY GR LEVEL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS	3410.572	1	3410.572	185.311	0.000
LEVEL	3410.572	1	3410.572	185.311	0.000
EXPLAINED	3410.572	1	3410.572	185.311	0.000
RESIDUAL	1656.417	90	18.405		
TOTAL	5066.989	91	55.681		

92 CASES WERE PROCESSED.
0 CASES (0.0 PCT) WERE MISSING.

27 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - Bush ICL 2988 EMAS

GROUP 1 - LEVEL EQ 1.
GROUP 2 - LEVEL EQ 2

TABLE 34

VARIABLE		NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	F VALUE	2-TAIL PROB.	POOLED VARIANCE ESTIMATE T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.	SEPARATE VARIANCE ESTIMATE T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
GR													
	GROUP 1	47	44.4468	3.917	0.571	1.41	0.253	13.61	90	0.000	13.56	86.09	0.000
	GROUP 2	45	32.2667	4.649	0.693								
LIS													
	GROUP 1	47	8.8298	1.129	0.165	1.80	0.053	15.82	90	0.000	15.92	84.99	0.000
	GROUP 2	45	5.5333	0.842	0.126								
DIC1													
	GROUP 1	47	56.4681	4.091	0.597	2.92	0.000	13.67	90	0.000	13.53	70.33	0.000
	GROUP 2	45	40.2222	6.993	1.042								
DIC2													
	GROUP 1	47	61.8085	3.982	0.581	3.09	0.000	14.45	90	0.000	14.29	69.11	0.000
	GROUP 2	45	44.7333	7.005	1.044								
CL1													
	GROUP 1	47	6.7234	2.456	0.358	1.20	0.346	6.34	90	0.000	6.36	89.80	0.000
	GROUP 2	45	3.5111	2.247	0.334								

TABLE 35

CL2													
	GROUP 1	47	7.1489	1.829	0.267	1.33	0.341	5.77	90	0.000	5.75	87.02	0.000
	GROUP 2	45	4.7778	2.109	0.314								
CL3													
	GROUP 1	47	8.9362	1.799	0.262	2.38	0.004	6.96	90	0.000	6.90	74.90	0.000
	GROUP 2	45	5.9556	2.776	0.414								
CL4													
	GROUP 1	47	8.3617	1.762	0.257	2.30	0.006	8.42	90	0.000	8.35	75.67	0.000
	GROUP 2	45	4.4000	2.675	0.399								
CLT													
	GROUP 1	47	31.1277	5.900	0.861	1.71	0.074	9.02	90	0.000	8.97	82.34	0.000
	GROUP 2	45	18.2444	7.720	1.151								
SP													
	GROUP 1	47	4.2128	0.623	0.091	1.25	0.454	15.41	90	0.000	15.45	89.58	0.000
	GROUP 2	45	2.3111	0.957	0.083								
TEACT													
	GROUP 1	47	206.6170	14.595	2.129	2.09	0.015	16.76	90	0.000	16.63	77.93	0.000
	GROUP 2	45	143.4667	21.091	3.144								

TABLE 36

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
UR	0.4709	0.4086	0.4503	0.5955	0.2589	-0.3114	0.4248	0.1335	0.2141	0.2725
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.014	P=0.031	P=0.019	P=0.002	P=0.125	P=0.082	P=0.026	P=0.279	P=0.172	P=0.113
ABS	0.4001	0.1749	0.0139	0.1164	0.2680	0.2445	0.3712	0.1822	0.3547	0.5787
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.034	P=0.221	P=0.476	P=0.305	P=0.117	P=0.139	P=0.046	P=0.211	P=0.055	P=0.003
UL	0.5679	0.5316	0.3066	0.2112	0.4880	-0.2021	0.4573	0.2245	0.3753	0.2772
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.003	P=0.006	P=0.085	P=0.176	P=0.011	P=0.186	P=0.017	P=0.160	P=0.044	P=0.109

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPLTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
UR	0.6178
	(22)
	P=0.001
ABS	0.3825
	(22)
	P=0.041
UL	0.5314
	(22)
	P=0.006

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 37

13 MAR 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
UR ↓	0.6948	0.6559	0.6698	0.6708	0.4247	0.3000	0.5310	0.4481	0.6077	0.4937
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.000	P=0.001	P=0.000	P=0.000	P=0.026	P=0.090	P=0.006	P=0.019	P=0.001	P=0.010
ABS	0.7041	0.6782	0.3190	0.2683	0.7070	0.3318	0.6132	0.5371	0.7780	0.8422
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.000	P=0.000	P=0.076	P=0.116	P=0.000	P=0.068	P=0.001	P=0.005	P=0.000	P=0.000
UL	0.7569	0.5919	0.5248	0.4169	0.2780	0.4003	0.6071	0.3699	0.5939	0.6434
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.000	P=0.002	P=0.007	P=0.028	P=0.108	P=0.034	P=0.002	P=0.047	P=0.002	P=0.001

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

13 MAR 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
UR	0.7709
	(22)
	P=0.000
ABS	0.6342
	(22)
	P=0.001
UL	0.6573
	(22)
	P=0.000

TABLE 38

13 MAR 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
UR	0.6186	0.4174	0.2333	0.3541	0.7311	0.6767	0.9231	0.4801	0.7900	0.5164
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P=0.057	P=0.161	P=0.296	P=0.204	P=0.023	P=0.037	P=0.001	P=0.123	P=0.012	P=0.103
ABS	0.4460	0.2258	0.0696	0.4529	0.5570	0.5083	0.4775	0.0304	0.4160	0.7184
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P=0.143	P=0.303	P=0.438	P=0.139	P=0.083	P=0.107	P=0.124	P=0.473	P=0.162	P=0.026
UL	0.8726	0.7467	0.9271	0.4849	0.1529	0.4358	0.3118	0.3194	0.3567	0.7559
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P=0.003	P=0.019	P=0.001	P=0.120	P=0.364	P=0.149	P=0.235	P=0.229	P=0.202	P=0.017

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

13 MAR 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edirburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
UR	0.7766
	(8)
	P=0.014
ABS	0.4780
	(8)
	P=0.124
UL	0.7693
	(8)
	P=0.015

TABLE 39

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
UR	0.6494	0.4560	0.1370	0.2946	0.4819	0.4296	0.6506	0.5825	0.6388	0.6364
	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
	P=0.003	P=0.035	P=0.303	P=0.129	P=0.027	P=0.045	P=0.003	P=0.008	P=0.004	P=0.003
ABS	0.6468	0.4643	0.2328	0.4229	0.2521	0.2935	0.3386	0.1307	0.2933	0.5689
	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
	P=0.003	P=0.032	P=0.188	P=0.048	P=0.168	P=0.130	P=0.095	P=0.311	P=0.130	P=0.009
UL	0.5862	0.4526	0.1894	0.1955	0.3426	0.3285	0.3778	0.5276	0.4617	0.8246
	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
	P=0.007	P=0.036	P=0.237	P=0.230	P=0.092	P=0.102	P=0.070	P=0.016	P=0.033	P=0.000

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
UR	0.6320
	(17)
	P=0.004
ABS	0.5451
	(17)
	P=0.013
UL	0.5177
	(17)
	P=0.018

TABLE 40

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GP	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
UR	0.5942	0.6628	0.6767	0.6901	0.2766	-0.2019	0.4633	0.2751	0.3339	0.4910
	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
	P=0.011	P=0.004	P=0.003	P=0.002	P=0.164	P=0.239	P=0.044	P=0.165	P=0.116	P=0.034
ABS	0.5581	0.3863	0.3092	0.1986	0.2030	0.2576	0.4314	0.1520	0.3588	0.5528
	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
	P=0.017	P=0.081	P=0.135	P=0.243	P=0.238	P=0.181	P=0.057	P=0.298	P=0.098	P=0.018
UL	0.4802	0.4118	0.4793	0.5021	0.5268	-0.4110	0.4264	0.1164	0.3239	0.2887
	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
	P=0.037	P=0.067	P=0.038	P=0.030	P=0.024	P=0.067	P=0.060	P=0.345	P=0.124	P=0.153

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
UR	0.7597
	(15)
	P=0.001
ABS	0.5211
	(15)
	P=0.025
UL	0.5807
	(15)
	P=0.013

TABLE 41

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
ECT	0.5713	0.5766	0.4278	0.3439	0.4354	0.5406	0.6970	0.4805	0.7101	0.6328
	(47)	(47)	(47)	(47)	(47)	(47)	(47)	(47)	(47)	(47)
	P=0.000	P=0.000	P=0.001	P=0.009	P=0.001	P=0.000	P=0.000	P=0.000	P=0.000	P=0.000

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
ECT	0.7286
	(47)
	P=0.000

- - - - - - - - - - - - PEARSON CORRELATION COEFFICIENTS - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
ECT	0.5316	0.4891	0.3666	0.4261	0.1742	0.3702	0.4610	0.3995	0.4559	0.3191
	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)	(45)
	P=0.000	P=0.000	P=0.007	P=0.002	P=0.126	P=0.006	P=0.001	P=0.003	P=0.001	P=0.016

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - PEARSON CORRELATION COEFFICIENTS - - - - - - - - - - - -

	TEACT
ECT	0.5782
	(45)
	P=0.000

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

PRECEDING TASK REQUIRED 1.00 SECONDS CPU TIME; 27.00 SECONDS ELAPSED.

13 0 FINISH

TABLE 43

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
RATINGT	0.6705	0.5378	0.1458	0.2818	0.3291	0.3108	0.4978	0.4927	0.4810	0.7000
	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
	P=0.002	P=0.014	P=0.291	P=0.140	P=0.102	P=0.116	P=0.022	P=0.024	P=0.027	P=0.001

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
RATINGT †	0.5749
	(17)
	P=0.009

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 44

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
RATINGT	0.8299	0.7259	0.4472	0.5363	0.2335	0.4237	0.5822	0.2928	0.4457	0.5774
/	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P=0.006	P=0.024	P=0.142	P=0.093	P=0.296	P=0.157	P=0.071	P=0.249	P=0.143	P=0.074

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
RATINGT	0.7374
/	(8)
	P=0.021

PEARSON CORRELATION COEFFICIENTS

	RATINGT	UL	ABS	UR
RATINGT	1.0000 (0) P=*****	0.9613 (45) P=0.000	0.6291 (45) P=0.000	0.8207 (45) P=0.000
UL	0.9613 (45) P=0.000	1.0000 (0) P=*****	0.5941 (45) P=0.000	0.7758 (45) P=0.000
ABS	0.6291 (45) P=0.000	0.5941 (45) P=0.000	1.0000 (0) P=*****	0.6897 (45) P=0.000
UR	0.8207 (45) P=0.000	0.7758 (45) P=0.000	0.6897 (45) P=0.000	1.0000 (0) P=*****

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 46

- - - - - - - - - - - - - PEARSON CORRELATION COEFFICIENTS - - - - - - - - - -

	RATINGT	UL	ABS	UR
RATINGT	1.0000 (0) P=*****	0.7376 (47) P=0.000	0.5030 (47) P=0.000	0.8660 (47) P=0.000
UL	0.7376 (47) P=0.000	1.0000 (0) P=*****	0.3940 (47) P=0.003	0.6387 (47) P=0.000
ABS	0.5030 (47) P=0.000	0.3940 (47) P=0.003	1.0000 (0) P=*****	0.4356 (47) P=0.001
UR	0.8660 (47) P=0.000	0.6387 (47) P=0.000	0.4356 (47) P=0.001	1.0000 (0) P=*****

TABLE 47

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
RATINGT	0.7557	0.8137	0.6679	0.3765	0.5175	-0.0872	0.5842	0.3088	0.5205	0.4082
	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
	P=0.001	P=0.000	P=0.004	P=0.087	P=0.026	P=0.381	P=0.012	P=0.136	P=0.025	P=0.069

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
RATINGT	0.8213
	(15)
	P=0.000

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 48

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
RATINGT	0.7438	0.5477	0.5338	0.4832	0.3397	0.4964	0.6842	0.3181	0.6593	0.6900
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.000	P=0.005	P=0.006	P=0.012	P=0.063	P=0.010	P=0.000	P=0.077	P=0.000	P=0.000

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
RATINGT	0.7061
	(22)
	P=0.000

TABLE 49

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
RATINGT	0.7643	0.7368	0.4580	0.1713	0.4732	-0.0047	0.5655	0.3103	0.4857	0.3627
	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)	(22)
	P=0.000	P=0.000	P=0.017	P=0.226	P=0.014	P=0.492	P=0.003	P=0.082	P=0.012	P=0.050

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
RATINGT	0.7367
	(22)
	P=0.000

TABLE 50

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

PRECEDING TASK REQUIRED 1.00 SECONDS CPU TIME; 23.00 SECONDS ELAPSED.

```
  7  0          temporary
  8  0          select if (class=2)
  9  0          pearson corr
 10  0           math to ft with gr to teact
```

*****PEARSON CORR PROBLEM REQUIRES 4752 BYTES WORKSPACE *****

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	OR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
MATH	0.3210 (13) P=0.148	0.2598 (13) P=0.201	0.3444 (13) P=0.130	0.4148 (13) P=0.084	0.4908 (13) P=0.047	0.3296 (13) P=0.141	0.2865 (13) P=0.176	0.6261 (13) P=0.012	0.6229 (13) P=0.013	0.5512 (13) P=0.028
PHY	0.5985 (13) P=0.017	0.5685 (13) P=0.023	0.6004 (13) P=0.017	0.6520 (13) P=0.009	0.3349 (13) P=0.137	0.5781 (13) P=0.021	0.3952 (13) P=0.095	0.7885 (13) P=0.001	0.7579 (13) P=0.002	0.7547 (13) P=0.002
DYN	0.5317 (13) P=0.033	0.5595 (13) P=0.025	0.4635 (13) P=0.059	0.5019 (13) P=0.043	0.5844 (13) P=0.020	0.1973 (13) P=0.264	0.1709 (13) P=0.293	0.0433 (13) P=0.445	0.3523 (13) P=0.124	0.5073 (13) P=0.041
AE	0.3466 (13) P=0.128	0.3328 (13) P=0.138	0.3025 (13) P=0.163	0.3849 (13) P=0.102	0.6401 (13) P=0.010	0.0716 (13) P=0.410	-0.0009 (13) P=0.499	0.0198 (13) P=0.475	0.2627 (13) P=0.198	0.4326 (13) P=0.074
AVN	0.6454 (13) P=0.010	0.6368 (13) P=0.011	0.6685 (13) P=0.007	0.7227 (13) P=0.003	0.6052 (13) P=0.016	0.4408 (13) P=0.070	-0.1113 (13) P=0.362	0.4724 (13) P=0.055	0.5304 (13) P=0.034	0.7784 (13) P=0.001
AIR	0.4523 (13) P=0.064	0.4596 (13) P=0.061	0.3513 (13) P=0.124	0.3884 (13) P=0.099	0.5603 (13) P=0.025	0.1064 (13) P=0.368	0.1945 (13) P=0.267	-0.0076 (13) P=0.490	0.2963 (13) P=0.168	0.3834 (13) P=0.103
NAV	0.3982 (13) P=0.093	0.4116 (13) P=0.085	0.2945 (13) P=0.170	0.3403 (13) P=0.133	0.5647 (13) P=0.024	0.0068 (13) P=0.491	0.0270 (13) P=0.464	-0.1087 (13) P=0.365	0.1710 (13) P=0.292	0.3279 (13) P=0.142
MET	0.4752 (13) P=0.054	0.4744 (13) P=0.054	0.4090 (13) P=0.087	0.4571 (13) P=0.062	0.6712 (13) P=0.007	0.1219 (13) P=0.349	0.0446 (13) P=0.444	0.0453 (13) P=0.443	0.3165 (13) P=0.151	0.4668 (13) P=0.057
FT	0.5076 (13) P=0.041	0.4865 (13) P=0.049	0.5131 (13) P=0.039	0.5986 (13) P=0.017	0.5305 (13) P=0.034	0.4291 (13) P=0.076	0.0632 (13) P=0.421	0.3875 (13) P=0.100	0.5199 (13) P=0.037	0.7203 (13) P=0.003

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 51

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME PAGE
University of Edinburgh - ERCC ICL 2976 EMAS

PRECEDING TASK REQUIRED 1.00 SECONDS CPU TIME; 23.00 SECONDS ELAPSED.

```
  11  0          temporary
  12  0          select if (class=3)
  13  0          pearson corr
  14  0           math to ft with gr to teact
```

*****PEARSON CORR PROBLEM REQUIRES 4752 BYTES WORKSPACE *****

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME PAGE
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - PEARSON CORRELATION COEFFICIENTS - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
MATH	0.4239 (12) P=0.090	0.3222 (12) P=0.159	-0.0652 (12) P=0.422	0.0987 (12) P=0.383	0.2376 (12) P=0.234	-0.0220 (12) P=0.474	0.4048 (12) P=0.101	0.3951 (12) P=0.107	0.3548 (12) P=0.134	0.2984 (12) P=0.179
PHY	0.1726 (12) P=0.300	0.0593 (12) P=0.429	-0.1280 (12) P=0.349	0.0087 (12) P=0.490	0.1042 (12) P=0.377	-0.1031 (12) P=0.378	0.1535 (12) P=0.321	0.1540 (12) P=0.320	0.1177 (12) P=0.361	0.0437 (12) P=0.448
DYN	0.1175 (12) P=0.361	0.0347 (12) P=0.458	0.1063 (12) P=0.374	-0.1668 (12) P=0.306	0.1144 (12) P=0.365	0.3643 (12) P=0.128	-0.1104 (12) P=0.369	0.1311 (12) P=0.346	0.1357 (12) P=0.319	0.1805 (12) P=0.292
AE	0.0974 (12) P=0.384	0.0760 (12) P=0.409	0.0793 (12) P=0.406	-0.2275 (12) P=0.244	-0.1766 (12) P=0.296	0.1995 (12) P=0.272	0.0314 (12) P=0.462	0.2821 (12) P=0.193	0.1039 (12) P=0.377	0.1962 (12) P=0.275
AVN	0.2716 (12) P=0.202	0.2226 (12) P=0.249	0.0906 (12) P=0.392	-0.1453 (12) P=0.330	0.1165 (12) P=0.362	0.2191 (12) P=0.252	0.3657 (12) P=0.127	0.2047 (12) P=0.267	0.2833 (12) P=0.192	0.3878 (12) P=0.112
AIR	0.3911 (12) P=0.110	0.3918 (12) P=0.109	0.0628 (12) P=0.425	-0.0360 (12) P=0.457	0.0541 (12) P=0.435	0.2079 (12) P=0.263	0.2811 (12) P=0.194	0.6632 (12) P=0.011	0.4144 (12) P=0.095	0.3479 (12) P=0.139
NAV	0.6265 (12) P=0.016	0.6005 (12) P=0.021	0.2140 (12) P=0.257	0.3918 (12) P=0.109	0.1984 (12) P=0.273	0.2228 (12) P=0.248	0.4000 (12) P=0.104	0.7386 (12) P=0.003	0.5361 (12) P=0.039	0.4869 (12) P=0.058
MET	0.4492 (12) P=0.076	0.3956 (12) P=0.107	0.2077 (12) P=0.263	0.0699 (12) P=0.417	0.0474 (12) P=0.443	0.4285 (12) P=0.087	0.3763 (12) P=0.119	0.6060 (12) P=0.020	0.4716 (12) P=0.065	0.4401 (12) P=0.081
FT	0.6850 (12) P=0.008	0.6336 (12) P=0.015	0.2303 (12) P=0.241	0.1687 (12) P=0.304	0.7077 (12) P=0.006	0.4083 (12) P=0.099	0.5084 (12) P=0.049	0.4627 (12) P=0.069	0.6967 (12) P=0.007	0.7011 (12) P=0.006

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 52

```
17 JUL 86    SPSS-X RELEASE 1.0 A  ICL 2900  EMAS/VME
             University of Edinburgh - ERCC ICL 2976        EMAS

0- - - - - - - - - - - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S - - - - - - - - - - - - -

                GR          TEACT       CLT

 FM              0.7447      0.5914      0.7860
                 (    5)     (    5)     (    5)
                 P=0.086     P=0.162     P=0.067

 PHY             0.7913      0.6492      0.7532
                 (    5)     (    5)     (    5)
                 P=0.065     P=0.132     P=0.082

 NAV             0.8011      0.7643      0.6823
                 (    5)     (    5)     (    5)
                 P=0.061     P=0.077     P=0.116

 MET             0.3990      0.2984      0.2844
                 (    5)     (    5)     (    5)
                 P=0.268     P=0.325     P=0.333

 FT              0.4998      0.4330      0.4967
                 (    5)     (    5)     (    5)
                 P=0.212     P=0.249     P=0.213

 AV              0.7644      0.6300      0.7390
                 (    5)     (    5)     (    5)
                 P=0.077     P=0.142     P=0.088

0(COEFFICIENT / (CASES) / SIGNIFICANCE)      (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)
0
17 JUL 86    SPSS-X RELEASE 1.0 A  ICL 2900  EMAS/VME
             University of Edinburgh - ERCC ICL 2976.        EMAS

0PRECEDING TASK REQUIRED        1.00 SECONDS CPU TIME;      15.00 SECONDS ELAPSED.

0       4 COMMAND LINES READ.
        0 ERRORS DETECTED.
        0 WARNINGS ISSUED.
        1 SECONDS CPU TIME.
       31 SECONDS ELAPSED TIME.
          END OF JOB.
```

TABLE 53

24 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - -

	TEACT
MATH	0.3927 (39) P=0.007
PHY	0.3321 (39) P=0.019
DYN	0.3871 (39) P=0.007
AE	0.2908 (39) P=0.036
AVN	0.1144 (39) P=0.244
AIR	0.3250 (39) P=0.022
NAV	0.3307 (39) P=0.020
MET	0.4306 (39) P=0.003
FT	0.5194 (39) P=0.000

TABLE 54

----PEARSON CORRELATION COEFFICIENTS----------

	QR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
MATH	0.3583 (39) P=0.013	0.3002 (39) P=0.032	0.2767 (39) P=0.044	0.2959 (39) P=0.034	0.3971 (39) P=0.006	0.2459 (39) P=0.066	0.4573 (39) P=0.002	0.5577 (39) P=0.000	0.5403 (39) P=0.000	0.3888 (39) P=0.007
PHY	0.2903 (39) P=0.036	0.2260 (39) P=0.083	0.2372 (39) P=0.073	0.2902 (39) P=0.037	0.3115 (39) P=0.027	0.3078 (39) P=0.028	0.3314 (39) P=0.020	0.4511 (39) P=0.002	0.4499 (39) P=0.002	0.3313 (39) P=0.020
DYN	0.4188 (39) P=0.004	0.3990 (39) P=0.006	0.3451 (39) P=0.015	0.3848 (39) P=0.008	0.2696 (39) P=0.048	0.2202 (39) P=0.089	0.0730 (39) P=0.329	0.0803 (39) P=0.314	0.2054 (39) P=0.105	0.4161 (39) P=0.004
AE	0.3162 (39) P=0.025	0.2822 (39) P=0.041	0.2332 (39) P=0.076	0.2905 (39) P=0.036	0.2469 (39) P=0.065	0.1100 (39) P=0.253	0.0363 (39) P=0.413	0.1273 (39) P=0.220	0.1735 (39) P=0.145	0.3776 (39) P=0.009
AVN	0.2019 (39) P=0.109	0.2144 (39) P=0.095	-0.0672 (39) P=0.342	0.0505 (39) P=0.380	0.3935 (39) P=0.007	0.2298 (39) P=0.080	0.1065 (39) P=0.259	0.1030 (39) P=0.266	0.2711 (39) P=0.047	0.3942 (39) P=0.006
AIR	0.3882 (39) P=0.007	0.3827 (39) P=0.008	0.2672 (39) P=0.050	0.2968 (39) P=0.033	0.2065 (39) P=0.104	0.1120 (39) P=0.249	0.0997 (39) P=0.273	0.0972 (39) P=0.278	0.1684 (39) P=0.153	0.3452 (39) P=0.016
NAV	0.4154 (39) P=0.004	0.4034 (39) P=0.005	0.2411 (39) P=0.070	0.3226 (39) P=0.023	0.2233 (39) P=0.086	0.0524 (39) P=0.376	0.0917 (39) P=0.289	0.1421 (39) P=0.194	0.1721 (39) P=0.147	0.3677 (39) P=0.011
MET	0.4633 (39) P=0.001	0.4282 (39) P=0.003	0.3695 (39) P=0.010	0.4042 (39) P=0.005	0.2481 (39) P=0.064	0.1846 (39) P=0.130	0.1373 (39) P=0.202	0.2508 (39) P=0.062	0.2667 (39) P=0.050	0.4439 (39) P=0.002
FT	0.4558 (39) P=0.002	0.4329 (39) P=0.003	0.5490 (39) P=0.000	0.5085 (39) P=0.000	0.1164 (39) P=0.240	0.2027 (39) P=0.108	0.2536 (39) P=0.060	0.3605 (39) P=0.012	0.2965 (39) P=0.033	0.4581 (39) P=0.002

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

PRECEDING TASK REQUIRED 0.00 SECONDS CPU TIME; 15.00 SECONDS ELAPSED.

```
  7  0          temporary
  8  0          select if (class=2)
  9  0          pearson corr
 10  0           gpa with gr to teact
```

*****PEARSON CORR PROBLEM REQUIRES 528 BYTES WORKSPACE *****

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
GPA	0.5381	0.5360	0.4745	0.5390	0.6722	0.2138	0.1157	0.1576	0.4161	0.5748
	(13)	(13)	(13)	(13)	(13)	(13)	(13)	(13)	(13)	(13)
	P=0.031	P=0.032	P=0.054	P=0.031	P=0.007	P=0.246	P=0.356	P=0.307	P=0.083	P=0.022

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
GPA	0.5343
	(13)
	P=0.032

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

PRECEDING TASK REQUIRED 1.00 SECONDS CPU TIME; 17.00 SECONDS ELAPSED.

```
  11  0        temporary
  12  0        select if (class=3)
  13  0        pearson corr
  14  0         gpa with gr to teact
```

*****PEARSON CORR PROBLEM REQUIRES 528 BYTES WORKSPACE *****

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	GR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
GPA	0.4299	0.3581	0.0529	0.0335	0.1885	0.1733	0.3528	0.4704	0.4005	0.3829
	(12)	(12)	(12)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
	P=0.086	P=0.132	P=0.437	P=0.460	P=0.283	P=0.299	P=0.136	P=0.065	P=0.104	P=0.115

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
GPA	0.3039
	(12)
	P=0.174

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	OR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
MATH	0.5801 (14) P=0.016	0.4837 (14) P=0.043	0.5817 (14) P=0.016	0.5421 (14) P=0.025	0.6102 (14) P=0.011	0.4076 (14) P=0.078	0.5769 (14) P=0.017	0.6616 (14) P=0.006	0.6926 (14) P=0.003	0.5448 (14) P=0.024
PHY	0.3797 (14) P=0.094	0.2657 (14) P=0.184	0.2891 (14) P=0.163	0.3696 (14) P=0.101	0.4089 (14) P=0.077	0.3237 (14) P=0.134	0.3348 (14) P=0.126	0.5111 (14) P=0.033	0.4824 (14) P=0.043	0.3699 (14) P=0.101
DYN	0.5931 (14) P=0.014	0.5564 (14) P=0.021	0.5518 (14) P=0.022	0.6096 (14) P=0.011	0.5983 (14) P=0.013	0.6337 (14) P=0.008	0.5899 (14) P=0.014	0.5458 (14) P=0.024	0.7063 (14) P=0.003	0.6322 (14) P=0.008
AE	0.5179 (14) P=0.031	0.4348 (14) P=0.064	0.3992 (14) P=0.083	0.3789 (14) P=0.095	0.4478 (14) P=0.057	0.3685 (14) P=0.102	0.4894 (14) P=0.041	0.6402 (14) P=0.008	0.5972 (14) P=0.013	0.5099 (14) P=0.034
AVN	0.2505 (14) P=0.199	0.4116 (14) P=0.076	0.2187 (14) P=0.231	0.2434 (14) P=0.206	0.1660 (14) P=0.289	0.1239 (14) P=0.340	0.4125 (14) P=0.075	0.3690 (14) P=0.101	0.3352 (14) P=0.125	0.4344 (14) P=0.064
AIR	0.1783 (14) P=0.275	0.2352 (14) P=0.214	0.2236 (14) P=0.226	0.1317 (14) P=0.330	0.3590 (14) P=0.108	0.2923 (14) P=0.160	0.3244 (14) P=0.133	0.0497 (14) P=0.434	0.3006 (14) P=0.153	0.2364 (14) P=0.213
NAV	0.2480 (14) P=0.201	0.2281 (14) P=0.221	0.1044 (14) P=0.364	0.1616 (14) P=0.294	0.2987 (14) P=0.155	0.1678 (14) P=0.287	0.3611 (14) P=0.107	0.2807 (14) P=0.170	0.3418 (14) P=0.120	0.3544 (14) P=0.111
MET	0.5277 (14) P=0.028	0.4402 (14) P=0.061	0.3825 (14) P=0.093	0.4632 (14) P=0.051	0.4715 (14) P=0.047	0.3578 (14) P=0.109	0.4805 (14) P=0.044	0.5546 (14) P=0.022	0.5706 (14) P=0.018	0.5277 (14) P=0.028
FT	0.5420 (14) P=0.025	0.5185 (14) P=0.031	0.3671 (14) P=0.103	0.4214 (14) P=0.070	0.4723 (14) P=0.047	0.2540 (14) P=0.195	0.5583 (14) P=0.021	0.4353 (14) P=0.063	0.5309 (14) P=0.027	0.6047 (14) P=0.012

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

TABLE 58

26 FEB 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

- - - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - - -

	TEACT
MATH	0.2490 (12) P=0.223
PHY	0.0534 (12) P=0.436
DYN	0.1219 (12) P=0.356
AE	0.0833 (12) P=0.401
AVN	0.2278 (12) P=0.243
AIR	0.2751 (12) P=0.199
NAV	0.4885 (12) P=0.057
MET	0.3688 (12) P=0.124
FT	0.5671 (12) P=0.030

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 59

----------PEARSON CORRELATION COEFFICIENTS----

	TEACT
MATH	0.4160 (13) P=0.083
PHY	0.6664 (13) P=0.007
DYN	0.5030 (13) P=0.043
AE	0.3569 (13) P=0.120
AVN	0.6956 (13) P=0.005
AIR	0.4119 (13) P=0.085
NAV	0.3472 (13) P=0.127
MET	0.4598 (13) P=0.061
FT	0.5548 (13) P=0.027

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

PEARSON CORRELATION COEFFICIENTS

	TEACT
MATH	0.6184 (14) P=0.010
PHY	0.3869 (14) P=0.090
DYN	0.6308 (14) P=0.009
AE	0.4721 (14) P=0.047
AVN	0.2686 (14) P=0.181
AIR	0.2189 (14) P=0.231
NAV	0.2239 (14) P=0.225
MET	0.4940 (14) P=0.039
FT	0.4787 (14) P=0.044

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

TABLE 61

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	QR	LIS	DIC1	DIC2	CL1	CL2	CL3	CL4	CLT	SP
GPA	0.5703 (14) P=0.018	0.5190 (14) P=0.031	0.4565 (14) P=0.054	0.4891 (14) P=0.041	0.5712 (14) P=0.018	0.4194 (14) P=0.071	0.6130 (14) P=0.011	0.6144 (14) P=0.011	0.6783 (14) P=0.004	0.6165 (14) P=0.010

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

- - - - - - - - - - - - P E A R S O N C O R R E L A T I O N C O E F F I C I E N T S - - - - - - - - - - - -

	TEACT
GPA	0.5621 (14) P=0.020

TABLE 62

017 JUL 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

PEARSON CORRELATION COEFFICIENTS

	GR	TEACT	CLT
MATH	0.6480 (12) P=0.013	0.5614 (12) P=0.031	0.5840 (12) P=0.025
PHY	0.4883 (12) P=0.057	0.3806 (12) P=0.116	-0.4680 (12) P=0.067
DYN	0.5416 (12) P=0.037	0.4457 (12) P=0.078	0.4755 (12) P=0.063
AE	0.5233 (12) P=0.044	0.4816 (12) P=0.060	0.6247 (12) P=0.017
AVN	0.7409 (12) P=0.003	0.6800 (12) P=0.008	0.6670 (12) P=0.010
AIR	0.6787 (12) P=0.009	0.7478 (12) P=0.003	0.7203 (12) P=0.005
NAV	0.4558 (12) P=0.073	0.5944 (12) P=0.023	0.4124 (12) P=0.096
MET	0.7421 (12) P=0.003	0.7319 (12) P=0.004	0.6968 (12) P=0.007
FT	0.8109 (12) P=0.001	0.7429 (12) P=0.003	0.7640 (12) P=0.002
AV	0.7148 (12) P=0.005	0.6118 (12) P=0.019	0.5185 (12) P=0.045

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT

017 JUL 86 SPSS-X RELEASE 1.0 A ICL 2900 EMAS/VME
University of Edinburgh - ERCC ICL 2976 EMAS

PEARSON CORRELATION COEFFICIENTS

	GR	TEACT	CLT
MATH	0.5484 (22) P=0.004	0.5065 (22) P=0.009	0.4695 (22) P=0.015
PHY	0.3982 (22) P=0.035	0.3673 (22) P=0.048	0.3627 (22) P=0.051
DYN	0.5199 (22) P=0.007	0.5102 (22) P=0.008	0.4842 (22) P=0.012
AE	0.4768 (22) P=0.013	0.4725 (22) P=0.014	0.5160 (22) P=0.008
AVN	0.6898 (22) P=0.000	0.6733 (22) P=0.000	0.6393 (22) P=0.001
AIR	0.4837 (22) P=0.012	0.4275 (22) P=0.025	0.4731 (22) P=0.014
NAV	0.2412 (22) P=0.143	0.2454 (22) P=0.138	0.1711 (22) P=0.226
MET	0.5706 (22) P=0.003	0.6130 (22) P=0.001	0.5614 (22) P=0.004
FT	0.7339 (22) P=0.000	0.7271 (22) P=0.000	0.7288 (22) P=0.000
AV	0.5434 (22) P=0.005	0.5288 (22) P=0.006	0.4516 (22) P=0.018

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT

TABLE 64

```
017 JUL 86   SPSS-X RELEASE 1.0 A  ICL 2900  EMAS/VME
             University of Edinburgh - ERCC ICL 2976          EMAS

- - - - - - - - - - - - - - P E A R S O N   C O R R E L A T I O N   C O E F F I C I E N T S

                GR              TEACT           CLT

MATH          0.2534          0.4235          0.3688
              (  10)          (  10)          (  10)
              P=0.246         P=0.118         P=0.154

PHY           0.0770          0.3173          0.2642
              (  10)          (  10)          (  10)
              P=0.419         P=0.193         P=0.237

DYN           0.4657          0.6547          0.6207
              (  10)          (  10)          (  10)
              P=0.093         P=0.022         P=0.031

AE            0.3669          0.4517          0.4428
              (  10)          (  10)          (  10)
              P=0.155         P=0.101         P=0.106

AVN           0.5410          0.6640          0.6337
              (  10)          (  10)          (  10)
              P=0.058         P=0.020         P=0.027

AIR           0.3655          0.2918          0.4275
              (  10)          (  10)          (  10)
              P=0.156         P=0.213         P=0.115

NAV          -0.0447         -0.0197          0.0151
              (  10)          (  10)          (  10)
              P=0.453         P=0.479         P=0.484

MET           0.3523          0.5526          0.5185
              (  10)          (  10)          (  10)
              P=0.166         P=0.053         P=0.067

FT            0.5419          0.7050          0.7205
              (  10)          (  10)          (  10)
              P=0.057         P=0.013         P=0.011

AV            0.2580          0.4492          0.4174
              (  10)          (  10)          (  10)
              P=0.242         P=0.103         P=0.122

(COEFFICIENT / (CASES) / SIGNIFICANCE)      (A VALUE OF 99.0000 IS PRINTED IF A COEFFI
```

بسم الله الرحمن الرحيم

145.

KING FAISAL AIR ACADEMY
ACADEMIC WING

END OF INTERMEDIATE EXAMINATION RESULT

COURSE - 32A

S. NO.	SER. NO.	NAME	MATH 100	PHY 100	DYN 100	A/E 100	AVN 100	AIR 100	NAV 100	MET 100	I/S 100	F/T 100	L/R 100	ADM 100	MGY 100	TOTAL 1300	%AGE	O/M
1.	1639	Hatem Bakht Abdallah Al-Zahrani	92	82	92	93	95	95	95	94	76	87	91	89	75	1156	88.92	1
2.	1493	Muhammad Hussain Ali Al-Anazi	89	87	91	83	95	88	96	88	81	67	91	99	92	1147	88.23	2
3.	1481	Abdulkareem Saud Radhi Al-Deham	85	87	92	90	80	94	95	96	83	77	94	85	82	1140	87.69	3
4.	1631	Abdallah Saleh Sa'eed Al-Ghamdi	85	84	90	84	94	92	97	93	69	83	77	70	82	1100	84.62	4
5.	1569	Hazaa Nasser Hazaa Al-Shahwani (Qatari)	76	83	92	83	94	89	97	96	60*	88	83	65	73	1079	83.00	5
6.	1344	Saeed Ali Yahya Al-Qahtani	79	75	89	83	94	87	88	77	71	63	91	92	84	1073	82.54	6
7.	1467	Eisa Hasan Mfarrih Asseri	74	83	96	85	95	85	89	92	64	77	60	84	78	1062	81.69	7
8.	1466	Ali Mbarak Muhammad Al-Rabi'ey	89	87	94	87	89	90	89	88	60	66	81	60*	61	1041	80.08	8
9.	1478	Thabit Abdallah Faiz Al-Amri	62	67	87	77	90	89	67	80	73	73	78	83	78	1004	77.23	9
10.	1347	Abdallah Muslem Rashed Al-Joheni	73	81	80	77	88	70	87	77	61	60*	85	87	72	998	76.77	10
11.	1377	Abdallah Awadh Sa'eed Al-Yahya	61	74	85	74	92	87	83	78	68	60*	77	82	67	988	76.00	11
12.	1498	Abdallah Ghanim Abdallah Hadhram	62	63	84	77	91	90	74	77	84	68	75	60	77	982	75.54	12
13.	1447	Muhammad Rafae Muhammad Al-Amri	62	69	81	73	93	89	92	75	60*	69	82	60*	67	972	74.77	13
14.	1583	Falah Ibdah Shabeeb Al-Khalidi	78	79	84	70	82	82	88	78	62	62	60	75	71	971	74.69	14

* Passed in Re-sit.

(MOHAMMED NASSER AL-HAZER)
Major,
Commander Academic Wing, KFAA

Dated:- 30th October.1985

KING FAISAL AIR ACADEMY
ACADEMIC WING
END OF INTERMEDIATE EXAMINATION RESULT

COURSE - 32C

S. NO.	SER. NO.	NAME	MATH 100	PHY 100	DYN 100	A/E 100	AVN 100	AIR 100	NAV 100	MET 100	I/S 100	F/T 100	L/R 100	ADM 100	MGY 100	TOTAL 1300	%AGE	O/M
1.	1558	Sulaiman Abdulrahman Muhammad Al-Wabel	95	98	95	96	97	92	94	99	82	90	89	70	77	1174	90.31	1
2.	1581	Jazi Ghazi Hamood Al-Otaibi	87	89	96	96	92	90	85	96	88	79	93	92	82	1165	89.62	2
3.	1599	Ali Awadh Dhafer Qahtani	89	94	96	97	96	88	94	92	88	84	82	92	72	1164	89.54	3
4.	1545	Ahmad Abdulrahman Yousif Al-Salem	91	92	95	96	90	91	94	94	82	81	86	80	76	1148	88.31	4
5.	1551	Ali Abdulrahman Ali Menbet	85	85	89	91	92	86	94	93	74	83	95	86	65	1118	86.00	5
6.	1598	Khalid Ahmad Muhammad Al-Halaiw	87	87	87	82	92	87	88	86	77	87	87	88	66	1101	84.69	6
7.	1555	Khalid Hamdan Hassan Al-Zahrani	90	95	96	60*	87	92	86	89	69	69	83	88	64	1068	82.15	7
8.	1568	Ali Bayat Muhammad Al-Assiry (Qatari)	75	75	92	92	89	87	88	93	73	81	60	83	63	1051	80.85	8
9.	1590	Muhammad Ali Muhammad Aal-Hashan	63	81	89	60*	91	83	85	80	93	72	82	82	82	1043	80.23	9
10.	1484	Ali Saleh Ali Al-Qazlan	76	77	83	75	66	89	93	73	73	63	91	85	60	1004	77.23	10
11.	1500	Hussain Hamed Muhammad Al-Raqeeb	79	79	70	66	78	85	85	80	71	60*	88	88	61	990	76.15	11
12.	1560	Saleh Muhammad Hadi Al-Qahtani	78	84	86	68	80	86	69	77	66	73	85	67	60*	979	75.31	12
13.	1491	Aazeh Dhaifallah Bin Ulayyan Al-Harbi	65	60*	79	67	75	80	87	80	69	60**	60*	81	70	933	71.77	13
14.	1386	Khalid Saleh Muhammad Al-Ghamdi	76	82	-	-	72	-	-	-	66	60	72	73	62	DISCHARGED		

* Passed in Re-sit.
** Passed in Re-resit

Dated:- 14th January, 1986

(MOHAMMED NASSER AL-HAZER)
Major,
Commander Academic Wing, KFAA

ACADEMIC WING

END OF INTERMEDIATE EXAMINATION RESULT

COURSE - 32D

S. NO.	SER. NO.	NAME	MATH 100	PHY 100	DYN 100	A/E 100	AVN 100	AIR 100	NAV 100	MET 100	I/S 100	F/T 100	L/R 100	ADM 100	MGY 100	TOTAL 1300	%AGE	O/M
1.	(2) 1549	Hamood Muhammad Fayez Al-Asmari	94	91	94	83	80	94	92	99	90	97	95	96	76	1181	90.85	1
2.	1669	Nasser Hijab Abdallah Al-Harbi	98	96	96	86	75	91	88	95	89	88	97	94	67	1160	89.23	2
3.	1528	Dhafir Abdulrahman Abdallah Al-Shehri	80	78	78	85	76	91	95	93	87	84	92	86	72	1097	84.39	3
4.	1691	Ali Sulaiman Abdallah Al-Amro	81	79	93	83	79	87	84	93	84	93	86	88	62	1092	84.00	4
5.	(4) 1663	Muhammad Mrawwih Dhaher Al-Rwaili	80	80	85	70	60	87	96	90	89	89	96	83	65	1070	82.31	5
6.	1542	Abdallah Sulaiman Muhammad Al-Rethayy'a	74	76	87	76	67	81	86	84	74	87	89	95	73	1049	80.69	6
7.	1617	Yahya Abdallah Yahya Al-Aseeri	80	80	83	75	60*	85	91	95	73	83	91	77	64	1037	79.77	7
8.	1567	Shabbab Faisal Zayed Al-Subay'ee	79	86	84	78	60*	90	87	83	76	81	83	88	60	1035	79.62	8
9.	1618	Tali' Fahad Tali' Al-Mutairi	93	90	76	65	60*	76	90	75	82	87	94	86	60	1034	79.54	9
10.	1529	Umair Abdallah Saleh Al-Qahtani	84	81	81	72	73	81	79	91	78	88	77	86	61	1032	79.39	10
11.	1531	Hayef Nwaidees Hakem Al-Shammari	73	64	68	65	60*	83	80	77	76	90	61	89	74	960	73.85	11
12.	1544	Ali Muhammad A'idh Al-Qarni	69	67	80	81	51	94	87	87	78	95	95	71	71	Academy Council decision awaited		
13.	1446	Sa'eed Ali Awdah Al-Shehrani	44	60*	71	60	55	59	55	55	60	79	84	73	62	-do-		

*Passed in Re-sit.

Dated:- 16th February, 1986

(MUHAMMAD NASSER AL-HAZER)
Lt. Col.
Commander Academic Wing

KING FAISAL AIR ACADEMY
ACADEMIC WING

END OF INTERMEDIATE EXAMINATION RESULT

COURSE - 33A

S. NO.	SER. NO.	NAME	MATH 100	PHY 100	DYN 100	A/E 100	AVN 100	AIR 100	NAV 100	MET 100	F/T 100	GUR 100	I/S 100	ADM 100	MGY 100	L/R 100	TOTAL 1400	%AGE	O/M
1.	1701	Elwalid Abdel Salam Elmahboub Moham ed (Sudani)	99	96	98	87	91	94	94	97	94	91	86	92	83	87	1289	92.07	1
2.	1571	Ali Saeed Abdallah Al-Ghamdi	95	96	96	86	87	95	86	97	84	91	92	85	69	94	1253	89.50	2
3.	1592	Ibraheem Yahya Hussain Al-Qahtani	96	91	96	85	83	94	93	91	76	76	85	90	77	95	1228	87.71	3
4.	1700	Hamid Ahmed Gar-Elnabi (Sudani)	97	97	92	75	81	87	90	95	91	84	83	82	80	92	1226	87.57	4
5.	1714	Abdelrahman Bushara Ahmed Bushara (Sudani)	94	94	87	77	81	88	85	93	81	85	86	84	85	87	1207	86.21	5
6.	1512	Dahman Ali Dahman Al-Shehri	86	84	82	64	71	87	90	89	78	92	93	95	90	94	1195	85.36	6
7.	1543	Abdulrahman Muhammad Hassan Aseeri	88	89	93	75	83	90	84	92	73	77	83	84	86	93	1190	85.00	7
8.	1711	Salah Mohammed Khalifa Al-Ammari (Bahraini)	97	97	90	94	89	94	92	96	96	65	60	60	60	75	1165	83.21	8
9.	1707	Khalid Muham ad Abdallah Al-atteyah (Qatari)	83	90	91	75	75	94	95	88	74	68	63	84	65	60*	1105	78.93	9
10.	1548	Saeed Ali Nasser Al-Ghamdi	77	82	70	68	60	89	98	87	61	86	76	79	68	87	1084	77.43	10
11.	1559	Fahad Shafi Aidh Al-Qahtani	80	72	86	68	63	91	92	76	60*	82	76	64	62	84	1056	75.43	11
12.	1519	Muhammad Ali Saeed Aal-Murait	76	81	87	80	60	90	83	82	60*	76	67	65	60	86	1053	75.21	12
13.	1525	Misleh Mbarak Faleh Al-Dosari	77	82	79	65	60	81	84	64	60*	73	65	62	62	76	990	70.71	13
14.	1372	Faleh Mohsen Mnahi Al-Ajmi	68	69	76	63	63	91	93	77	67	62	60*	60	62	72	983	70.21	14

* Passed in Re-sit.

Dated:- 16th March, 1986

(MOHAMMAD NASSER AL-HAZER)
Lt.Col.
Commander Academic Wing, KFAA

KING FAISAL AIR ACADEMY
ACADEMIC WING

END OF INTERMEDIATE EXAMINATION RESULT

COURSE - 33B

S. NO.	SER. NO.	NAME	MATH 100	PHY 100	DYN 100	A/E 100	AVN 100	AIR 100	NAV 100	MET 100	F/T 100	OUR 100	I/S 100	ADM 100	MGY 100	L/R 100	TOTAL 1400	%AGE	O/M
1.	1702	Mohamed Sa'ad Mohamed Hilmi (Sudani)	97	97	95	89	88	94	96	96	97	77	89	92	98	93	1298	92.71	1
2.	1677	Muhammad Hussain Ahmad Aal-Naji	92	89	91	94	95	97	93	96	86	81	89	92	78	91	1264	90.29	2
3.	1626	Abdulrahman Muhammad Ali Aal-Mizher	88	85	92	87	78	91	91	96	90	85	76	96	85	90	1230	87.86	3
4.	1637	Bader Abdrabbeh Raddah Al-Malki	95	94	90	89	74	96	91	94	88	91	84	90	61	79	1216	86.86	4
5.	1684	Farhan Amer Farhan Al-Shehri	89	80	76	76	75	87	97	78	60*	94	91	90	77	91	1161	82.93	5
6.	1624	Khalid Saud Abdulrahman Al-Khasham	74	78	76	84	79	83	91	83	65	76	77	88	82	80	1116	79.71	6
7.	1673	Aidh Awadh Talib Al-Rasheedi	87	82	78	79	79	82	81	81	71	88	60	86	73	77	1104	78.86	7
8.	1609	Adnan Ahmad Muhammad Al-Melhem	82	81	78	69	60	90	84	89	74	66	67	79	74	74	1067	76.21	8
9.	1577	Abdallah Fay'a Abdallah Aal-Awwadh	75	74	74	60	73	82	81	69	71	74	63	79	86	80	1041	74.36	9
10.	1616	Muhammad Matooq Ali Al-Malki	83	83	61	60*	60	90	78	64	60*	90	77	77	61	96	1040	74.29	10
11.	1611	Muhammad Awadhallah Khadher Al-Thebaiti	76	79	70	60*	68	87	71	82	66	75	67	70	76	89	1036	74.00	11
12.	1586	Sulaiman Ali Muhammad Al-Fayez	74	74	72	63	65	67	98	72	64	74	64	86	66	92	1031	73.64	12
13.	1708	Abdullah Nasser Abdullah Al-Mesned (Qatari)	76	72	73	74	71	93	86	79	71	61	60*	82	61	68	1027	73.36	13
14.	1645	Abdel-I'Lah Ibraheem Nasser Al-Fa'iz	63	63	47	41	60	72	90	44	53	67	62	78	77	84	DECISION AWAITED		

*Passed in Re-sit.

Dated:- 19th April, 1986

(MOHAMMED NASSER AL-HAZER)
Lt.Col
Commander Academic Wing, KFAA

KING FAISAL AIR ACADEMY
ACADEMIC WING

INTERMEDIATE PHASE PROGRESS TEST RESULT

COURSE - 33C

S. NO.	SER. NO.	NAME	F/M 100	PHY 100	NAV 100	MET 100	F/T 100	TOTAL 500	%AGE	O/M
1.	1739	Khalid Homoud Shafaq Al-Dhimairi	88	98	100	88	94	468	93.6	1
2.	1688	Shara' Ali Mbarak Jerman	86	96	100	92	91	465	93.0	2
3.	1602	Suhail Metlaq Mansour Al-Otaibi	82	82	100	88	89	441	88.2	3
4.	1540	Fahad Abdulrahman Ibraheem Al-Hraish	80	97	100	79	82	438	87.6	4
5.	1544	Ali Muhammad A'idh Al-Qarni	74	77	100	93	93	437	87.4	5
6.	1518	Ahmad Abdallah Hazmi Al-Ahmari	68	89	83	95	88	423	84.6	6
7.	1634	Abdul Majeed Al-Shareef Mastoor Al-Fi'er	56	83	100	79	74	392	78.4	7
8.	1550	Ibraheem Yahya Hazza'a Aseeri	75	86	95	63	66	385	77.0	8
9.	1479	Hassan Sa'eed Jamah Al-Shehri	66	69	73	84	87	379	75.8	9
10.	1710	Thabit Muhammad Jabar Al-Dosari (Bahraini)	60	78	89	71	77	375	75.0	10
11.	1713	Khaled Abdulla Ali Al-Khalifa (Bahraini)	68	75	84	60	61	348	69.6	11
12.	1648	Faisal Ahmad Saleh Al-Sha'ia	42	71	93	72	69	347	69.4	12
13.	1709	Salman Abdulrahman Salman Al-Khalifa (Bahraini)	50	65	62	79	83	339	67.8	13
14.	1638	Abdallah Saleh Mar'ee Al-Qahtani	60	60	81	39	60	300	60.0	14
15.	1516	Sultan Shar Muhammad Al-Shehri	28	53	81	60	63	285	57.0	15

Dated:- 16th June, 198

(MOHAMMED AL-NASSER AL-HAZER)
Lt.Col.
Commander Academic Wing KFAA